Chronic Acromioclavicular Joint and Subacromial Pain: Reliability and Diagnostic Accuracy of Clinical Tests

Monique Baigent

2023

Health & Environmental Sciences, Physiotherapy

A thesis submitted to

Auckland University of Technology

in fulfilment of the requirements for the degree of Masters of Philosophy (MPhil)

Supervisors

Dr Steven G. White

Dr Julia Hill

Associate Professor Alain C. Vandal

Abstract

Shoulder pathologies are the third most common musculoskeletal complaint globally. In New Zealand they are responsible for substantial disability-adjusted life years and healthcare expenditure. Chronic shoulder pathologies are often misdiagnosed due to the poor validity of tests and an over-reliance on imaging. A correct diagnosis of shoulder pain is integral to forming a prognosis and management plan.

The aim of this thesis was to investigate the validity of commonly used clinical tests for identifying acromioclavicular joint (ACJ) and subacromial pain in patients with chronic shoulder pathologies.

A systematic review was carried out to investigate the reliability and diagnostic accuracy of orthopaedic special tests (OSTs) for ACJ and subacromial pain. Estimates of inter-rater reliability were poor to moderate for the majority of tests and the outcomes between the studies varied considerably. The diagnostic accuracy of OSTs was also poor, with no test consistently demonstrating adequate diagnostic accuracy to be a valid tool in clinical practice. When combined, clusters of clinical tests, other than OSTs, have shown adequate diagnostic accuracy to influence the probability of an ACJ or subacromial diagnosis.

A prospective inter-rater reliability study was completed with 20 participants presenting with shoulder pain. A standardised physical assessment was conducted twice for each participant by two blinded clinicians. The test results of each clinician were compared to calculate percentage agreement, prevalence-adjusted and bias-adjusted kappa (PABAK) and kappa scores. The outcomes demonstrated poor to moderate agreement for most tests (≤0.60 PABAK). The exception to this, with high inter-rater reliability, was observed muscle wasting, bilateral ACJ deformity, pseudoparalysis and the horizontal adduction with external rotation test (Add/ER).

A diagnostic accuracy study was conducted with 38 participants with ACJ or subacromial pain. Each participant underwent a standardised interview and physical assessment followed by anaesthetic injection into the subacromial space +/- the ACJ. A positive anaesthetic response threshold was set at ≥65% improvement on the numeric pain rating scale. None of the clinical variables reached statistical significance for

diagnostic accuracy, suggesting that they cannot be used as stand-alone tools to either rule in or out ACJ or subacromial pain in this cohort. A multivariate analysis identified a combination of 10 positive and negative variables that have the potential to be clinically useful to predict a subacromial diagnosis. These were: difficulty with overhead tasks, a strain injury onset, lowest pain ≥5/10, presence of muscle wasting, onset of pain from a repetitive activity, worst pain ≥8/10, the primary pain site over or above the clavicle, painful Add/ER, painful passive internal rotation and painful resisted flexion at 10 degrees. The area under the receiver operating characteristic curve for these variables was 0.73, with a misclassification error of 35%.

This thesis's findings support the current research that clinical variables as stand-alone tests are not able to accurately identify an ACJ or subacromial diagnosis. It is recommended that clinicians place more emphasis on a combination of findings from the wider clinical assessment. Future research should focus on examining the validity of combinations of tests with anaesthetic blocks as a reference test.

Contents

List of Figures

List of Tables

List of Appendices

List of Abbreviations

+LR: Positive likelihood ratio

-LR: Negative likelihood ratio

1° pain: Primary pain source

2° pain: Secondary pain source

ACJ: Acromioclavicular joint

AI: Anaesthetic injection

CACJ: Chronic acromioclavicular joint

CI: Confidence interval

CMDHB: Counties Manukau district health board

CSI: Central sensitisation inventory

DALY: Disability-adjusted life years

ER: External rotation

HBB: Hand behind back

MMT: Manual muscle testing

MRI: Magnetic resonance imaging

MRA: Magnetic resonance arthrography

NAR: Negative anaesthetic response

NPRS: Numeric pain rating scale

NPV: Negative predictive value

OSC: Outpatient orthopaedic shoulder clinic

OSS: Oxford shoulder score

OST: Orthopaedic special test

PABAK: Prevalence-adjusted bias-adjusted kappa

PAR: Positive anaesthetic response

PGPM: Patient generated provocative movement

PPV: Positive predictive value

PRISMA: Preferred reporting items for systematic reviews and meta-analyses

QAREL: Quality appraisal of reliability studies

QUADAS: Quality assessment of diagnostic accuracy studies

YLD: Years lived with disability

Attestation of Authorship

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person (except where explicitly defined in the acknowledgements), nor material which to a substantial extent has been submitted for the award of any other degree or diploma of a university or other institution of higher learning.

Signature	20/08/2023 Date

Research Outputs from this Thesis

National Conference Presentations:

Baigent, M., White, S., Hill, J., Vandal, A, C., Zo, M., & Chan, B. (2022). Are orthopaedic special tests obsolete in the diagnosis of acromioclavicular and subacromial pathology? Invited conference presentation NZSES: InterContinental, Natadola Bay, Fiji.

Planned Conference Presentations:

Baigent, M., White, S., Hill, J., Vandal, A, C., Zo, M., & Chan, B. (2022). Are orthopaedic special tests obsolete in the diagnosis of acromioclavicular and subacromial pathology? Invited conference presentation NZMPA: Novotel, Rotorua, New Zealand.

Acknowledgements

I would like to thank my supervisory team of Steve White, Julia Hill and Alain Vandal for providing me with guidance and support, attending countless meetings and putting in the hard work to make this project happen. Especially my primary supervisor Steve White who put many hours into shaping the quality of my writing and methodology.

I would like to acknowledge the New Zealand Manipulative Physiotherapists Association (NZMPA) and the trustees of the NZMPA scholarship fund for awarding me $5340 to assist in the completion of this research project. This allowed me to complete this project and appropriately acknowledge the study participant's time.

Many thanks to Michael Zo, Mr Hogan Yeung and Mr Brendan Coleman who patiently helped me in the clinic to collect the participant data and followed my lengthy assessment sheets. Particularly Michael who was the guinea pig for the assessment protocols and my sounding board throughout.

Benson Chan was my research assistant who triaged potential participants, called to book and consent them, printed and sent out all the patient information documents, and then consented patients on the day. It was a long two years of following multiple processes in repetition, many thanks for sticking with it.

Thank you to Valerie Richardson, the charge nurse of the Orthopaedic Outpatient Clinic for your patience in allowing me to restructure my clinic multiple times and the administration staff especially Shellie Garner for working around my study clinics.

Many thanks to my parents who supported me through this journey of learning, multiple frustrations and setbacks with food, moral support, and patience. I wouldn't have made it through without you.

Ethics Approval and Study Registration

Ethical approval and study registration for the research conducted as part of this thesis included:

- Health and Disability Ethics Committees (HDEC) reference 20\STH\203 accepted on the 3rd of December 2020
- Mātauranga Māori Committee, Auckland University of Technology, accepted on the 21st of October 2020
- The Counties Manukau Health Research Registry reference #1330 accepted on the 2nd of February 2021
- Māori ethics approval was given by the Counties Manukau Māori Health Development Team on the 6th of August 2020.
- Australia New Zealand Clinical Trial Registry (ANZCTR) registration, reference ACTRN12621000348853 accepted on the 26th of March 2021
- Prospero systematic review registration, reference CRD42022329336 accepted on the 3rd of June 2022

Copies of the approval letters, patient consent, and information sheets along with relevant application details can be found in the Appendix.

Chapter 1 Introduction

1.1 The problem

Musculoskeletal conditions are the most common cause of physical disability and loss of function in the developed world (Institute for Health Metrics and Evaluation, 2019; World Health Organisation, 2021). Globally they affect approximately 1.3 billion people and they are the largest contributor to years lived with disability (YLD) worldwide. This equates to approximately 149 million or 17% of the total YLDs (Safiri et al., 2021; World Health Organisation, 2021). These figures and the health burden musculoskeletal conditions create are expected to increase exponentially over time due to the projected growth of our ageing population (Fayaz et al., 2016; World Health Organisation, 2021). Musculoskeletal conditions commonly lead to chronic pain and significant limitations in mobility. Those affected experience negative impacts on their quality of life and physical and mental wellbeing (Dahlhamer et al., 2018). Consequently, musculoskeletal conditions have a sizable effect on economies due to time off work and medical care (World Health Organisation, 2021).

In New Zealand the incidence and burdens of musculoskeletal disorders are also rising. A Ministry of Health (2016) report collated from national health surveys calculated that musculoskeletal disorders are responsible for 13% of all health loss (Ministry of Health, 2016). This was measured in terms of disability-adjusted life years (DALY) and health-adjusted life expectancy. Worksafe (2019) has published a report revealing that in New Zealand, musculoskeletal disorders amount to 20% of all the work-related DALYs annually. The health expenditure related to the medical costs of managing these conditions and time away from work amounts to 20.8% of New Zealand's total health expenditure on non-communicable diseases (Blakely et al., 2019). Blakely et al. (2019) calculated this to be approximately $36.8 billion New Zealand dollars on a yearly basis measured over seven years from a retrospective data analysis.

Shoulder pain significantly contributes to this health cost as the third most common musculoskeletal disorder (Bento et al., 2019; Blyth et al., 2019). Recorded incidence rates for developing a shoulder condition are approximately 2.5% of the population and the

annual point prevalence figures have been shown to range from 6.9% to 26% (Kennedy et al., 2006). The lifetime prevalence has been calculated to be as high as 66.7% in the general population (Bento et al., 2019; Struyf et al., 2016). This varies based on the age groups concerned; those aged between 45-64 years have been identified as the most at risk group, with an incidence of 17.3 per 1,000 person-years (Djade et al., 2020). Shoulder pain is unfortunately also a persistent condition for many, with only 50% of patients reaching a full recovery within the timeframe of six months (Struyf et al., 2016). Shoulder pain affects all age groups and can arise insidiously or occur following trauma, strain injuries or overuse activities (Cadogan, Laslett, Hing, McNair, & Coates, 2011). Physically demanding occupations and overhead or contact sports are linked with a higher risk of developing a shoulder condition (Djade et al., 2020). The three most common structures responsible for causing shoulder pain are the rotator cuff, glenohumeral joint and acromioclavicular joint (ACJ) (Cadogan, Laslett, Hing, McNair, & Coates, 2011).

This thesis will focus on chronic ACJ (CACJ) and subacromial pain which is defined as pain persisting at these locations longer than the expected healing time of three to six months (Treede et al., 2015).

Subacromial pain conditions make up approximately 44-65% of presenting shoulder complaints (Consigliere et al., 2018). This condition is also referred to as subacromial impingement or rotator cuff related pain. Subacromial pain describes symptoms arising from the soft tissue structures in the space between the humeral head, and the coracoacromial arch (Cools et al., 2008). Pain here can occur due to a number of different pathologies including tears or tendinopathy of the rotator cuff tendons, calcific tendinitis and subacromial bursitis (Cadogan et al., 2016).

A number of chronic conditions can affect the ACJ including osteoarthritis, distal clavicle osteolysis, ligament damage and post traumatic arthritis (Shaffer, 1999). ACJ osteoarthritis is the most common cause of CACJ pain, and makes up approximately 20% of patients presenting with shoulder complaints (Farrell et al., 2019).

An accurate diagnosis of shoulder pain enables the clinician to triage a patient and tailor their treatment plan appropriately (Cadogan, McNair, Laslett, & Hing, 2013; Hegedus et al., 2008). It can also limit unnecessary imaging, injections and delays in effective management (Deyo, 2002). A missed underlying ACJ pathology following rotator cuff repair surgery has been shown to be one of the main factors leading to poor outcomes, emphasising the importance of an early accurate diagnosis (O'Holleran et al., 2005; Park et al., 2004).

Subacromial pain is the most common and well researched shoulder condition, consequently it is often over-diagnosed in clinical practice (De Yang Tien & Tan, 2014). In contrast CACJ pain is frequently poorly diagnosed and overlooked, as many of its signs and symptoms overlap with subacromial pain and these conditions can occur concurrently (Cools et al., 2008). Traumatic or acute ACJ pain is more commonly recognised as there is often a specific mechanism of injury, swelling, bruising, point tenderness and/or a separation deformity to guide the clinician (Sirin et al., 2018).

There is a paucity of good quality research for diagnosing CACJ and the current clinically utilised tests for both conditions have limited validity (Cadogan, McNair, Laslett, & Hing, 2013; Hegedus et al., 2015; Javed et al., 2017). The accepted reference standard for the diagnosis of ACJ or subacromial pain is significant pain relief following a guided anaesthetic injection (AI) into the space or joint respectively (Cadogan, McNair, Laslett, & Hing, 2013; Malfair, 2008). This however is an expensive, invasive, and often difficult to access modality for patients (James et al., 2005). The current best practice for a clinical diagnosis of shoulder pain is therefore made on the basis of information obtained from the interview and physical examination of the patient, and from medical imaging (Shaffer, 1999).

The patient interview provides essential information to the process of diagnosing ACJ or subacromial pain (Cadogan et al., 2012; Powell & Huijbregts, 2006; Raynor & Kuhn, 2016). Very few studies however have questioned the diagnostic accuracy of interview answers (Cadogan, McNair, Laslett, & Hing, 2013). A patient's symptom location, nature of pain, mechanism of injury and aggravating factors can provide significant clues to the likely structure or pathology causing their symptoms (Cadogan, McNair, Laslett, & Hing, 2013; Lewis, 2016; Salamh & Lewis, 2020).

With respect to the value of information from the physical examination, there is a lack of diagnostic accuracy studies published for ACJ pain. The available research indicates that diagnostic tests for the ACJ have limited validity as stand-alone tools (Cadogan, McNair, Laslett, & Hing, 2013). The most recent systematic review by Krill et al. (2018) examined the accuracy of ACJ clinical tests. These authors excluded any studies below Level II evidence (case-control and cohort studies) resulting in only two studies being included. Krill et al. (2018) concluded that there is a paucity of high-level studies in this field and that based on current evidence, no combination of orthopaedic special tests (OSTs) had a meaningful impact on the post-test probability of ACJ pathology being present.

In contrast, there is a vast amount of literature published on the diagnostic accuracy of OSTs for subacromial pain. The majority of the research however is poor to moderate quality (Hegedus et al., 2015). Amongst the high-quality studies, very few have adequate statistical significance to support the use of OSTs in the clinical setting. The conclusion reached by these studies is that no OSTs have adequate diagnostic accuracy to differentiate subacromial pain or specific rotator cuff tears from other structures in the shoulder (Gismervik et al., 2017; Green et al., 2008; Hegedus et al., 2008; Lädermann, Collin, et al., 2021).

Reaching a differential diagnosis on the basis of medical imaging however is also a flawed practice (Shubin Stein et al., 2001). The commonly utilised diagnostic imaging for ACJ and subacromial pathology has a high prevalence of false positive findings (Cadogan, Laslett, Hing, McNair, & Coates, 2011). A study by Shubin Stein et al. (2001) assessed asymptomatic shoulders using magnetic resonance imaging (MRI) and found that 68% of the ≤30 age group had evidence of osteoarthritis of the ACJ and 93% of the >30 group. A high prevalence of asymptomatic rotator cuff tears is also evident on MRI, with positive findings in up to 50% of the population >60 and as high as 80% in those >80 years old (Milgrom et al., 1995; Sher et al., 1995). Full thickness tears can be identified in up to 25% of individuals in their sixties, and in >50% of eighty year olds (Edwards et al., 2016). Sub-deltoid/subacromial bursa abnormalities are also commonly seen on ultrasound in asymptomatic individuals (Hodgson et al., 2012). Additionally, no correlation exists between imaging results from magnetic resonance arthrography (MRA), x-ray or ultrasound and patients who have a positive

anaesthetic response (PAR) to an ACJ injection (Cadogan, Laslett, Hing, McNair, & Coates, 2011; Javed et al., 2017). Cadogan, Laslett, Hing, McNair and Coates (2011) have demonstrated that subacromial bursitis or partial thickness rotator cuff tears on ultrasound are not associated with a PAR to a subacromial injection. Hence, there is a risk that abnormal findings identified via the patient assessment and medical imaging will be over emphasised, despite the proven poor correlation between such findings and pain of ACJ or subacromial origin (Cadogan, McNair, Laslett, & Hing, 2013).

For a clinical test to be valid it requires sufficient diagnostic accuracy and inter-rater reliability (Fritz & Wainner, 2001; Šimundić, 2009). Poor inter-rater reliability has a direct influence on a test's validity and may be the reason a patient receives conflicting diagnoses (Lange et al., 2017; May et al., 2010). The reliability of many diagnostic tests for ACJ and subacromial pain has not been investigated (Lange et al., 2017). To be able to discuss ACJ and subacromial pain diagnostic tests both aspects of validity need to be considered. Hence, this research will also examine the reliability of the tests included in the diagnostic accuracy study.

1.2 Thesis aims

The primary aim of this thesis is:

1. To provide a clear understanding of the diagnostic accuracy of clinical tests for chronic acromioclavicular joint and subacromial pain.
2. To examine the inter-rater reliability of commonly used clinical tests for chronic acromioclavicular joint and subacromial pain.
3. To understand the value of combinations of test findings as well as individual tests.

The secondary aims are:

1. To examine the prevalence of acromioclavicular joint and subacromial pathologies in Māori and Pasifika populations.

1.3 Overview of the thesis

The first chapter of the thesis outlines the impact of shoulder pain on society and the importance of diagnosing specific shoulder pathologies. This chapter also includes an introduction to chronic ACJ and subacromial pain and explains the flaws in the current methods for diagnosing these conditions.

Chapter two aims to describe the most commonly employed statistical measures in diagnostic accuracy research. These measures will be referred to throughout this thesis and will be integral to the interpretation of the literature and the study outcomes. The purpose of this section is to therefore facilitate the understanding around the terminology of the diagnostic accuracy measures that will feature in the following chapters.

Chapter three will present a comprehensive and in-depth systematic review of the current literature surrounding the following topics. Firstly, the reliability of currently utilised orthopaedic shoulder tests (OSTs) will be examined and secondly the diagnostic accuracy of OSTs for identifying ACJ and subacromial pain will be analysed. This section will present a brief introduction, an outline of the methodology then a critique of the available research. These reviews will utilise the PRISMA, QUADAS and QAREL tools designed to critique the quality of the chosen studies.

Chapter four will contain the reliability study which investigates the inter-examiner reliability of 25 physical shoulder tests (OSTs) between two clinicians. These tests are commonly utilised clinical tests with the design of identifying an ACJ or subacromial pathology. This section will consist of an introduction to the study, an outline of the methodology and subsequent results and discussion.

Chapter five will contain the diagnostic accuracy study which investigates the common clinical assessment (interview questions and physical tests) for ACJ and subacromial pain and compares each variable to the reference test of a diagnostic anaesthetic injection. This chapter will follow a similar structure to chapter four with an introduction to the study, the methods, results and discussion.

Chapter six will summarize the key findings of this thesis and how this compares to the available research. The results will be discussed in context with current practice and how the outcomes of this study may inform/influence clinicians.

1.4 Significance of the research

The misinterpretation of poor diagnostic tests can lead to an overutilisation of imaging, ineffective corticosteroid injections, failed rehabilitation, more visits to medical practitioners, increased medications prescribed, more time off work and/or even surgical procedures (Deyo, 2002). An inaccurate diagnosis therefore has the potential to result in increased costs to the healthcare system, delays in treatment and potentially increased risk for the patient (Deyo, 2002). Patients without timely appropriate management can develop persistent pain and experience multiple negative consequences including poor mood, disturbed sleep, changes in cognitive processes, decreased quality of life, economic difficulty and a reduced overall health status (Dahlhamer et al., 2018; Fine, 2011). Persistent pain can also lead to central nervous system changes and sensitisation resulting in hyperalgesia, dysesthesia and often long term disability (Fine, 2011).

Improved confidence in the diagnostic accuracy of tests for ACJ and subacromial pathology will facilitate an early, accurate diagnosis and lead to an appropriate treatment pathway. The ability to make an informed diagnosis will not only assist future patients with shoulder pain, it will also benefit physiotherapists, doctors and orthopaedic surgeons who manage patients with such pathologies. Determining the validity of clinical tests can therefore reduce the health burden of persistent ACJ and subacromial pain and the healthcare costs associated with misdiagnosis.

Chapter 2 Diagnostic Accuracy Measures

Reaching a diagnosis is an integral part of a clinician's role and involves the use of a range of tests. The outcomes of these tests are used to provide information to aide clinical decision making, however, not all tests are equal in respect to the accuracy of their ability to predict the presence or absence of a specific diagnosis (Hoo et al., 2017).

Diagnostic accuracy is a term that describes a test's ability to identify the target condition or rule out competing/absent conditions (Šimundić, 2009). The potential of a test to discriminate a condition correctly is measured by various methods. It is important to understand the metrics used so that the diagnostic utility of a test can be appropriately interpreted. The section below provides a description of the most commonly reported statistical measures of diagnostic accuracy. These measures will be the basis of the statistical analysis in the subsequent studies.

2.1.1 Sensitivity and specificity

The two most widely reported and utilised measures of diagnostic accuracy are sensitivity and specificity. Sensitivity is the average estimate of the percentage of people with the condition in question who receive a positive test i.e. 'true positives'. Whilst sensitivity is a measure of true positives, the real value of a very sensitive test is when that test is negative. If a test has 100% sensitivity, it will always be positive in people with the condition, hence there will never be any false negatives. Therefore, a negative result will enable the clinician to rule out the condition (Hegedus et al., 2015). The acronym SnNout helps to simplify how to interpret this measure (Power et al., 2013). If the test is sensitive (Sn) and the test result is negative (N), we can rule out (out) the condition of interest. However diagnostic tests are rarely 100% sensitive; if a test had a sensitivity of 75% for example, this means that there is a 25% chance of a false negative and incorrectly ruling out a condition.

Conversely, specificity is the average estimate of the portion of people correctly identified by the test as not having the diagnosis of concern i.e. 'true negatives' (Denegar & Fraser, 2006). Although specificity measures true negatives, if a test is very specific it is more valuable when the test is positive. For example, a test with 100% specificity it will always be negative in people without the condition, and accordingly there will be no false positives. A

positive result with a highly specific test therefore, will allow the clinician to rule in the condition with confidence (Hegedus et al., 2015). The acronym SPin is used to explain this measure (Power et al., 2013). If the test is specific (Sp) and the test result is positive (P), we can rule in (in) the condition of interest. A test with 75% specificity would have a 25% probability of a false positive and incorrectly ruling in a condition.

A sensitivity or specificity of at least 80-90% has been proposed in the literature to represent a robust measure of diagnostic accuracy (Power et al., 2013). Sensitivity and specificity are calculated from the data collected by diagnostic accuracy studies. Table 1 below, taken from Jaeschke et al. (1994), provides detail of how these findings are inputted into a 2x2 table and how these metrics are calculated from the data. The formula to calculate sensitivity is: true positive/(true positive + false negative), and the formula for specificity is: true negative/(true negative + false positive).

Table 1

Two by two table from Jaeschke et al. (1994)

Test result	**Reference standard**	
	Disease present	**Disease absent**
Disease present	True positive	False positive
Disease absent	False negative	True negative

Note. Reprinted from "Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group," by R. Jaeschke, G. H. Guyatt, and D. L. Sackett, 1994, *Journal of the American Medical Association, 271*(9), p. 706. Copyright 2015 by Elsevier B.V.

2.1.2 Predictive values

Predictive values can help a clinician to understand how a tests' outcome may influence their clinical decisions. They refer to the likelihood that a test has accurately identified the presence or absence of a condition once the test results are known (Safari et al., 2015).

In contrast to specificity and sensitivity, predictive values are affected by the condition's prevalence in the assessed population (Denegar & Fraser, 2006). When a condition is highly prevalent a test is more accurate at ruling in the condition and less accurate at ruling it out.

If a condition is rare, a test may have a low predictive value despite it being highly accurate (Shreffler & Huecker, 2023). Predictive values, therefore, are only useful when applied to the population in question and should not be extrapolated to groups with differences in disease prevalence (Šimundić, 2009).

A positive predictive value (PPV) describes the probability of the condition being present in a patient with a positive test. Conversely, negative predictive values (NPV) measure the probability of a patient not having the condition when a test is negative (Šimundić, 2009). A PPV of 85% indicates if the test is positive then the likelihood of the condition being present is 85%. The opposite is true for an NPV of 85%, which suggests that if a test is negative the chances of the condition being absent is 85%. When the prevalence of a condition increases the PPV also increases. Conversely as the prevalence decreases the NPV increases (Shreffler & Huecker, 2023).

2.1.3 Likelihood ratios

Likelihood ratios are derived from the sensitivity and specificity values. They are a more robust measure of diagnostic accuracy as they help the clinician to determine the size of any shift in probability of a condition being present or absent given the result of the test. A positive likelihood ratio (+LR) indicates, if a test is positive, the size of the shift in probability that the condition in question is present (Denegar & Fraser, 2006). A +LR is calculated using the following formula: +LR = Sensitivity/(1- Specificity) (Hegedus et al., 2015). A negative likelihood ratio (-LR) conversely signifies the probability that if a test result is negative, how likely it is that the condition is absent. A -LR is calculated with the following formula: -LR = (1 - Sensitivity)/Specificity (Denegar & Fraser, 2006). Table 2 below is from the Denegar and Fraser (2006) paper; it outlines the significance of the likelihood ratio values. A +LR of one can give a 50/50 chance of the condition being present. A +LR of five, however, can raise the probability of a condition being present by 30%. Furthermore, a +LR of 10 can increase the chances of a condition being ruled in by 45%. If there was already a prevalence of 50% for a condition in the population being tested, then a +LR of 10 provides a 95% chance of a true positive test (McGee, 2002). The reverse is true of 0.2 and 0.1 for a -LR; the chances of correctly ruling out a condition if the test is negative is increased by 30% if the -LR is 0.2 and 45% with 0.1.

Table 2
Approximate shift in probability associated with various likelihood ratios from Denegar & Fraser (2006)

Positive likelihood ratio	Negative likelihood ratio	Shift in probability condition is present
>10	<0.1	Large, often conclusive
5–10	0.1–0.2	Moderate but usually important
2–5	0.2–0.5	Small, sometimes important
1–2	0.5–1.0	Very small, usually unimportant

Note. Reprinted from "How useful are physical examination procedures? Understanding and applying likelihood ratios," by C. R. Denegar, and M, Fraser, 2006, *Journal of athletic training, 41*(2), p. 204. Copyright by the National Athletic Trainers' Association, Inc.

2.1.4 Pre and post-test probability

When examining a patient, the clinician will have a basis of suspicion and preconceived idea of a potential diagnosis. This is the pre-test probability; it occurs before any tests are done and it will vary depending on the patient in question. More specifically it depends on their demographics and the prevalence of the condition in different populations (Denegar & Fraser, 2006). The post-test probability is the chance of the diagnosis in question being present once the test has been completed and the result is known (Parikh et al., 2009). The post-test probability is based on the size of the likelihood ratio. It is calculated by the following equation: post-test odds/(post-test odds + 1). The post-test odds of a patient having the condition are based on the pre-test odds that the patient has the disease multiplied by the +LR.

2.1.5 P-values

A p-value measures the likelihood of supporting or rejecting the null hypothesis. The null hypothesis in a diagnostic accuracy study is there being no association between a test result and the diagnosis in question. The lower a p-value is, the more substantial the evidence that the null hypothesis can be rejected. Low p-values consequently give greater confidence or statistical significance to the alternative hypothesis. A statistically significant p-value is generally considered 0.05 or lower (Andrade, 2019; Schervish, 1996). Therefore, if the p-

value is ≤0.05 then you can reject the null hypothesis and be confident that there is a statistically significant relationship between the test results and the diagnosis.

2.1.6 Summary

The above statistics are all important measures when interpreting the accuracy of a test or group of tests. For example, if a test is highly specific, but the +LR is below two, or the p-value is higher than 0.05, the test still has not reached statistical significance. A working knowledge, therefore, of the basic diagnostic accuracy statistics can facilitate the correct interpretation of test results and ultimately aid in reaching a precise clinical diagnosis.

Chapter 3 The Validity of Orthopaedic Special Tests for Acromioclavicular Joint and Subacromial Pain: A Systematic Review

3.1 Introduction

The purpose of this chapter was to present the current literature on the reliability and diagnostic accuracy of orthopaedic special tests (OSTs) when diagnosing CACJ and subacromial pathologies. This systematic review serves as a foundation for the subsequent studies in the thesis.

OSTs are commonly used physical examination tools to diagnose or rule out a specific pathology. In the past they have been used under the assumption that they can distinguish between different pathologies of the rotator cuff tendons, ACJ, glenohumeral joint and/or bursa. This theory however is contradictory to the known anatomy of the shoulder, where many tissues are closely related and intertwined (Salamh & Lewis, 2020). This overlap of tissues has been demonstrated by electromyography showing the coactivation of multiple muscles during shoulder testing. Boettcher et al. (2009) demonstrated with ultrasonography that during the empty and full can tests (OSTs) at least eight other muscles were activated in addition to the supraspinatus, which the test was designed for.

Despite their widespread use, many systematic reviews and original articles have called into question the ability of OSTs to facilitate an accurate diagnosis (Hegedus et al., 2015). Much of the research available is poor to moderate quality, and very few articles produce statistically significant measures of diagnostic accuracy for OSTs as stand-alone tests. Hegedus et al. (2015), therefore, focussed on combinations of clinical findings from both the subjective and objective examination instead and identified three different groups of variables with strong likelihood ratios for diagnosing rotator cuff-related pain. The majority of the clinical variables identified in these groups, however, were not OSTs and instead measured muscle weakness, the outcomes from the patient interview, and demographic data. There is a large body of research examining the diagnostic accuracy of OSTs, however there is little research that has investigated the accuracy of information obtained from the patient interview or measures of impairment such as range of motion and strength testing

(Biederwolf, 2013; Hegedus et al., 2012). Further research is needed investigate these measures as stand-alone tests and when combined.

There are no current guidelines on the use of OSTs for medical practitioners to follow (Cadogan, Laslett, Hing, McNair, & Coates, 2011). A thorough medical guideline summarizing the best evidence-based practice for the use of OSTs would help to simplify and consolidate the assessment process. A clearer model for shoulder assessment may facilitate a more standardised diagnostic process. Misinterpretation of diagnostic tests and variations in test performance amongst clinicians are likely to lead to overutilisation of imaging, ineffective treatment, and more visits to medical practitioners (Deyo, 2002). The cost to the patient is likely to be a delayed recovery which has the potential to lead to persistent pain, economic difficulty, a reduced overall health status and in some cases long term disability (Dahlhamer et al., 2018; Fine, 2011; Friedman & Khan, 2019). The impact of misdiagnosis on the economy is increased imaging costs, health practitioner visits, days off work, medication costs and even surgery costs (Deyo, 2002). Reaching an accurate diagnosis is important as it can directly influence a patient's prognosis and treatment pathway, as well as decreasing costs to the healthcare system.

This systematic review aimed to update and critique the current literature concerning the reliability and diagnostic accuracy of OSTs, and the wider clinical assessment, for CACJ and subacromial pathologies. The objective was to determine if recent studies were reaching new conclusions regarding the accuracy of clinical tests. The latest systematic reviews addressing the validity of OSTs have been published more than two years ago and more current studies may provide additional evidence to influence their conclusions (Gismervik et al., 2017; Hegedus et al., 2015; Krill et al., 2018; Lädermann, Collin, et al., 2021; Lädermann, Meynard, et al., 2021; Schmidt et al., 2021). The two most current reviews have a narrow focus on a specific condition (Lädermann, Collin, et al., 2021; Lädermann, Meynard, et al., 2021); each review covers a different set of clinical variables and no current systematic reviews have been published that address both reliability and diagnostic accuracy. It was therefore considered valuable to conduct an update and critique of the most current systematic reviews.

The following questions were considered:

1. Do orthopaedic special tests for the shoulder have sufficient reliability to be used in a clinical setting?
2. Is there a consensus in the literature on the diagnostic accuracy of orthopaedic special tests for identifying chronic acromioclavicular joint and subacromial pathology?

3.2 Methods

This chapter identifies the literature that has been published on the validity of OSTs for ACJ and subacromial pain between January 1990 and October 2022. The eligibility criteria have been outlined, and research that met these criteria were investigated. The results of the search were presented, and the most current systematic reviews were identified for each of the following three sections: the reliability of OSTs, and the diagnostic accuracy of OSTs for ACJ and subacromial pain. These reviews were appraised using the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) tool; because they were high quality and the most recently published reviews, they were selected as a starting point to update the current research. Following selecting the reviews, any original studies published since were considered and critiqued in order to update the systematic reviews.

This systematic review followed the PRISMA guidelines to ensure transparency and quality of reporting (Page et al., 2021). The protocol including the full methodology and detailed search strategy were registered in advance with PROSPERO (CRD42022329336).

3.2.1 Identification and selection of the Literature

The original review was conducted in November 2019 to gather the literature and determine the gaps and opportunities for research. During the course of this study, further research had been published, therefore in 2022 the search was updated and completed. A computerised search was conducted in October 2022 using the following databases: MEDLINE, SPORTS Discus, CINAHL, PEDro, AMED, Cochrane Trials Register via Wiley and Scopus. The following key words were used for the diagnostic accuracy studies:

acromioclavicular OR ACJ OR subacromial OR "rotator cuff" OR "shoulder impingement" OR supraspinatus OR infraspinatus OR subscapularis OR "teres minor" OR shoulder AND diagnos* OR accura* AND specific* OR sensitiv* OR "likelihood ratio" OR "predictive value*" AND test* OR exam* OR "clinical test".

Key words utilised for the reliability study search were: acromioclavicular OR ACJ OR subacromial OR "rotator cuff" OR "shoulder impingement" OR supraspinatus OR infraspinatus OR subscapularis OR "teres minor" OR shoulder AND valid* OR reliable OR reliability OR inter-rater OR intra-rater AND kappa OR PABAK OR ICC OR "percentage agreement" OR ANOVA AND test* OR exam* OR "clinical test". Full details of search strategy are detailed in Appendices R and S.

3.2.2 Data sources and searches

The literature search identified several relevant systematic reviews; the intent was to update the most current reviews. If original articles had been published since the systematic review, an update of the recent evidence was undertaken (Garner et al., 2016). If no new research had been published, no further analysis was conducted.

3.2.3 Eligibility criteria

The inclusion criteria were prospective original journal articles and systematic reviews that examined the validity of OSTs for the diagnosis of ACJ or subacromial pain. Studies published in English between January 1990 to October 2022 were included. The restricted timeframe was pragmatic as only the most recent evidence was required to update the current systematic reviews. The population included adults aged 18 years or over with shoulder pain. Any reliability studies had to report on kappa, PABAK (prevalence-adjusted kappa values), intraclass correlation coefficient and/or percentage agreement. Diagnostic accuracy studies had to include at least one of the following: specificity, sensitivity, likelihood ratios and/or predictive values. The exclusion criteria were as follows: (1) studies that only examined tests for hypermobile shoulders, labral tears, scapula dysfunction, long head of biceps pain and/or stiff shoulders (2) studies that tested the validity of a device or machine.

3.2.4 Data extraction

Separate searches were conducted for the reliability and diagnostic accuracy studies. Studies were initially selected based on whether the title and/or abstracts fit the eligibility criteria. The next stage involved the selection of articles eligible for a full text review. The reasons for exclusion of studies at this stage are listed in the PRISMA flow diagram in the results section. Once relevant studies were selected, a citation search of these publications was conducted to identify further studies not found in the original search. Relevant reliability outcomes of percentage agreement, kappa and PABAK along with their stated confidence intervals (CIs) were recorded in an Excel spread sheet. The diagnostic accuracy measures extracted from the studies were specificity, sensitivity, likelihood ratios, diagnostic odds ratios, CIs, and positive and negative predictive values.

3.2.5 Quality assessment

The selected studies were then assessed using quality scoring tools. The relevant data to assess the quality of the study was obtained and extracted into table format. The potential for bias and quality of reporting was assessed across all studies.

The PRISMA tool was used to evaluate the quality of reporting in the systematic reviews. This tool is a validated list of 27 criteria which has been tailored for systematic reviews and meta-analyses. PRISMA was designed as a checklist for developing systematic reviews; however it can also be used as a tool to gauge quality for peer review and critical appraisal of research (Moher et al., 2009; Page et al., 2021). The PRISMA tool assesses for value, bias, and transparency (Page et al., 2021). The biases are reported in the discussion and the PRISMA scoring has been used as a reference point as follows: ≤15 indicated poor quality, 15.5 to 21 was moderate quality and between 21.5 to 27 was high quality (Li et al., 2014).

The original Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool was used to critique the diagnostic accuracy studies. The QUADAS has been validated as a reliable tool to assess the internal and external validity of diagnostic accuracy studies (Whiting et al., 2006). This tool includes 14 questions examining bias and study validity. If a study scored ≥8 on this scale (60%) it was considered to have a low risk of bias (Gismervik et al., 2017). The QUADAS

scores are outlined in table format in the results section and further reporting on bias can be found in the discussion.

Lastly the Quality Appraisal of Reliability studies (QAREL) checklist was used to assess the identified reliability studies. This is an 11-item quality assessment tool based on the QUADAS and STARD (Standards for Reporting Diagnostic accuracy studies) critiquing methods and adapted for reliability research (Lucas et al., 2013). Studies were critiqued on bias and those scoring <50% on the scale were considered poor quality, 50-66% moderate quality, and ≥67% high quality (Mani et al., 2017).

The lead researcher and a research assistant (MZ) independently assessed the selected articles and the systematic reviews using the above tools. The scores were compared, any disagreements were discussed, and a consensus was reached.

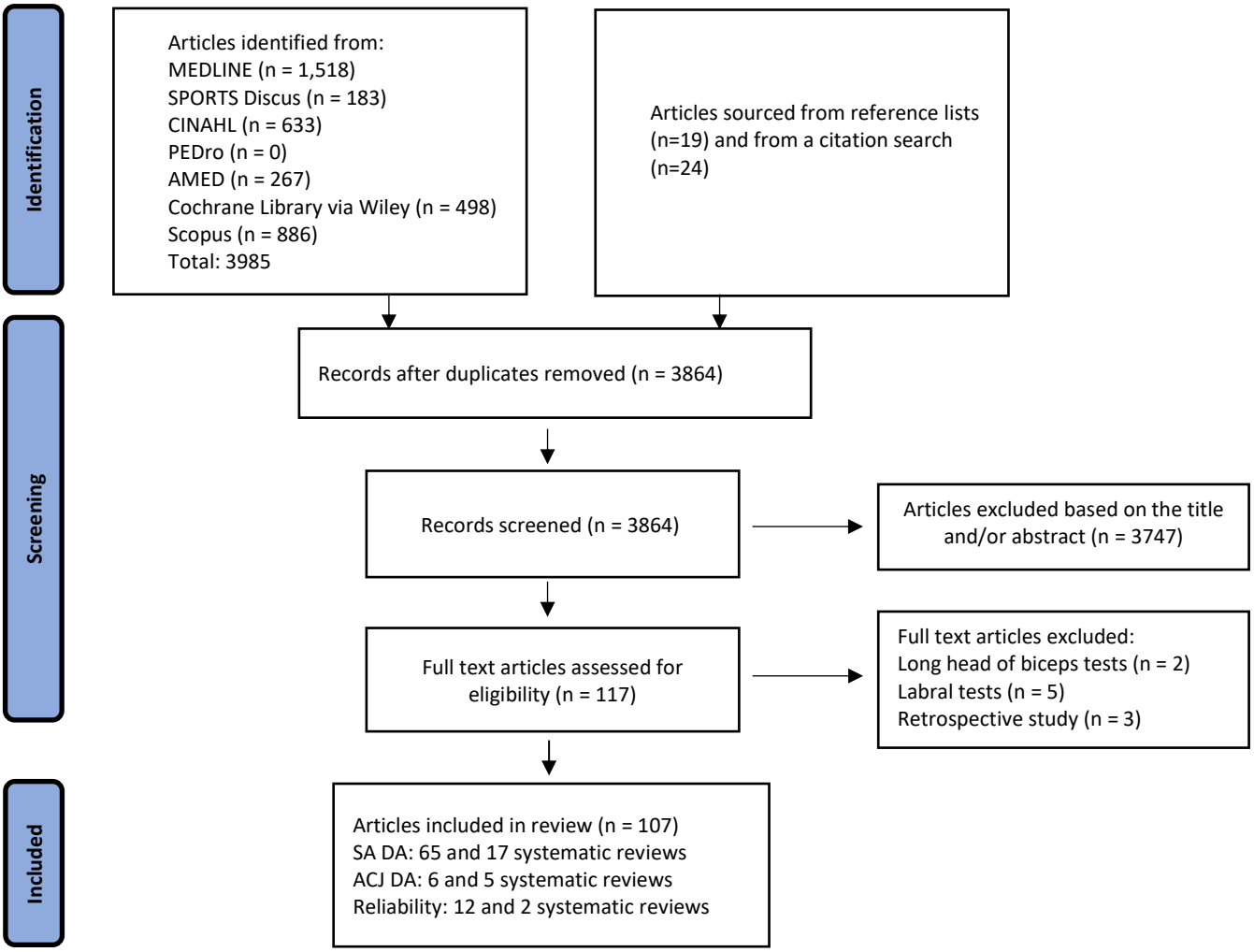
3.3 Results

3.3.1 Selection of studies

A summary of the literature search is provided below in Figure 1. In total, 3985 articles were identified in the initial search. They were screened using the inclusion/exclusion criteria, and 74 articles were selected. A search of the reference lists revealed 19 articles and 24 from the citations. Close examination of the full text articles eliminated a further 10 studies and the final number of relevant articles was 107.

Figure 1

PRISMA flow diagram for the validity of orthopaedic special tests for ACJ and SA pain



Note: SA: subacromial. DA: diagnostic accuracy. ACJ: acromioclavicular joint.

The search uncovered 65 original prospective articles and 17 systematic reviews on the diagnostic accuracy of subacromial tests. There were six original ACJ diagnostic accuracy studies identified and five systematic reviews. Twelve reliability studies were identified and two systematic reviews. Seven of these reliability studies assessed subacromial pain and the other five studies (Apeldoorn et al., 2021; Burns et al., 2016; Cadogan, Laslett, et al., 2011a; Nomden et al., 2009; Schmidt et al., 2021) and the systematic reviews examined the reliability of tests for general shoulder pain without a specific diagnosis. One original study by Michener et al. (2009), on subacromial pain, looked at both the diagnostic accuracy and

reliability of tests; and was therefore included in both groups. The included studies and their characteristics can be found in Appendix T.

3.3.2 Reliability of orthopaedic special tests

Twelve original reliability articles and two systematic reviews were retrieved that met the inclusion criteria. The most current, relevant systematic review was by Lange et al. (2017); it was therefore selected to critique and update. There were five original articles which have been published since the Lange et al. (2017) search which concluded on the 20th of March 2015.

This systematic review therefore, critiqued these five articles (Apeldoorn et al., 2021; Burns et al., 2016; Doxey et al., 2018; Ferenczi et al., 2018; Schmidt et al., 2021) and compared them to the Lange et al. (2017) review outcomes. See Table 3 for more detail of the study characteristics.

Table 3

Reliability study characteristics

Article	Design	No. of participants	Mean age	% of Females	Clinical tests
Apeldoorn et al., 2021	Inter-rater reliability study	113	38	55	Resisted ER, empty can, full can, active compression test, Neer Walsh, Hawkins-Kennedy, resisted IR, acromioclavicular joint stress.
Burns et al., 2016	Single-group repeated-measures interrater reliability study	21	41	67	Shoulder AROM & PROM. Resisted shoulder and elbow movements. Hawkins Kennedy, empty and full can tests, Neer Walsh, active compression test.
Doxey et al., 2018	Cross-sectional inter-rater reliability study	1070	42	66	Yergason, resisted ER, Neer Walsh, painful arc, Jobe test.
Ferenczi et al., 2018	Cross-sectional descriptive intra and inter-rater reliability study	34	60	26	Neer Walsh and Hawkins Kennedy.
Schmidt et al., 2021	Inter-rater reliability and diagnostic accuracy study	120	56	46	Belly press, ER lag sign, IR lag sign, painful arc, Hawkins Kennedy, resisted ER.
Lange et al., 2017	Reliability systematic review and meta-analysis	18 studies	NR	NR	Empty can, Neer Walsh, Hawkins Kennedy, painful arc.

Note: No.: number. ER: external rotation. IR: internal rotation. AROM: active range of motion. PROM: passive range of motion. NR: not recorded.

Article quality

The QAREL scores of the five articles are listed in Table 4. Apeldoorn et al. (2021), Ferenczi et al. (2018), Burns et al. (2016) and Schmidt et al. (2021) demonstrated moderate to high methodological quality. In these studies, the raters were blinded to the findings of other raters however, there was less clarity over blinding to additional information and the order of testing was not varied. The Doxey et al. (2018) reliability study in comparison, scored one point of nine on the QAREL scale. It had high levels of bias and poor methodology and therefore was not considered in the analysis.

Table 4

QAREL scoring of reliability articles

QAREL questions	Apeldoorn et al. (2021)	Burns et al. (2016)	Doxey et al. (2018)	Ferenczi et al. (2018)	Schmidt et al. (2021)
Subjects representative of the intended applied population	✓	✓	X	✓	✓
Raters representative of the intended applied population	✓	✓	✓	✓	✓
Raters blinded to the findings of other raters	✓	✓	X	✓	✓
Raters blinded to their own prior findings	N/A	N/A	N/A	?	N/A
Raters blinded to the results of the reference standard	N/A	N/A	N/A	N/A	✓
Raters blinded to information that was not intended to be part of the study	✓	?	X	?	?
Raters blinded to additional cues	✓	?	?	?	?
Varied order of examination	X	X	X	X	?
Appropriate timeframes between repeated measures	✓	✓	?	✓	?
Correct application and interpretation of tests	✓	✓	?	✓	✓
Appropriate statistical measures used	✓	✓	X	✓	✓
Total:	7/9	6/9	1/9	6/11	6/10

Note: N/A: not applicable. ✓: positive finding. X: negative finding.?: unclear finding.

Table 5 shows the PRISMA scoring of the included systematic reviews. The systematic review by Lange et al. (2017) demonstrated a 24/27 on the PRISMA scale demonstrating high-quality methodology and a good control of bias. The other four systematic reviews, which will be discussed below, all demonstrated moderate to high quality scores.

Table 4

PRISMA scoring of systematic reviews (reliability and diagnostic accuracy)

	Diagnostic Accuracy Reviews				Reliability Review
PRISMA Item	**Gismervik et al. (2017) [1]**	**Krill et al. (2018) [1]**	**Lädermann, Meynard, et al. (2021) [1]**	**Lädermann, Collin, et al. (2021) [1]**	**Lange et al. (2017) [2]**
1. Identified as a systematic review	✓	✓	X	✓	✓
2. Thorough abstract	X	X	X	X	X
3. Rationale for the review	✓	✓	✓	✓	✓
4. Objectives of the review	✓	✓	✓	✓	✓
5. Eligibility criteria	✓	✓	✓	✓	✓
6. Search tools used	✓	✓	✓	✓	✓
7. Search strategy	✓	✓	✓	✓	✓
8. Selection process	✓	✓	✓	✓	✓
9. Data collection process	✓	✓	✓	✓	✓
10. How data was sought	✓	✓	✓	✓	✓
11. Study risk of bias assessment	✓	✓	✓	✓	✓
12. Effect measures	✓	✓	✓	✓	✓
13. Synthesis methods	X	X	✓	✓	✓
14. Methods used to assess bias	✓	X	✓	✓	✓
15. Certainty assessment	✓	X	✓	✓	✓
16. Describe the results of the search	✓	✓	✓	✓	✓
17. Included study characteristics	✓	✓	✓	✓	X
18. Risk of bias in studies	✓	X	✓	✓	✓
19. Results of individual studies	✓	✓	✓	✓	✓
20. Results of syntheses of data	✓	✓	✓	✓	✓
21. Reporting bias	X	X	✓	✓	X
22. Certainty of evidence	✓	✓	✓	✓	✓
23. Clear, sequential discussion	✓	✓	✓	✓	✓
24. Registration and protocol	X	X	✓	✓	✓
25. Support (financial/non-financial)	✓	✓	✓	✓	✓
26. Conflict of interest	✓	✓	✓	✓	✓
27. Public availability of data	✓	X	✓	✓	✓
Total:	23/27	19/27	25/27	26/27	24/27

Note: ✓: positive finding; X: negative finding; ?: unclear finding; [1]: diagnostic accuracy review; [2]: reliability review.

The PRISMA abstract score was not attained by any of the reviews. This was due to the source of funding and registration numbers not being included in the abstract. All studies however met the other 10 of the 12 abstract criteria. The overall clarity and sequential ordering of information was of a high standard for these reviews.

Outcomes

The test outcomes from the reliability studies were varied in their results, Table 5 outlines the kappa, PABAK and agreement outcomes. The belly press test with a kappa of 0.63 (95% CI 0.25, 1.01), and the internal rotation lag sign with 1.00 (95% CI 1.00, 1.00) both demonstrated high values of inter-rater reliability (Schmidt et al., 2021). The Neer Walsh and empty can tests demonstrated moderate to high reliability, ranging from 0.43-0.64 and 0.51-0.69 kappa respectively (Apeldoorn et al., 2021; Burns et al., 2016; Lange et al., 2017). The Hawkins Kennedy test had the largest variance of kappa scores, from poor to high, ranging from 0.25 (95% CI 0.01, 0.48) up to 0.71 (95% CI 0.41, 1.00) (Burns et al., 2016; Schmidt et al., 2021). Resisted external rotation also demonstrated wide variability from the lowest kappa value of 0.22 (95% CI 0.00, 0.67) to a high value of 0.70 (95% CI 0.48, 0.92) (Burns et al., 2016; Schmidt et al., 2021). All other tests scored poor to moderate reliability. Burns et al. (2016) reported the second lowest kappa of 0.23 (95% CI 0.00, 0.65) for the full can test. Whilst the original studies only reported kappa, the systematic review by Lange et al. (2017) presented both PABAK and kappa values.

Table 5

Inter-rater reliability results of physical tests

Test	Author	Agreement (95% CI)	PABAK (95% CI)	Kappa (95% CI)
Active compression test (O'Brien test)	Apeldoorn et al. (2021)	0.74 (64.80, 81.70)	NR	0.46 (0.29, 0.63)
	Burns et al. (2016)	0.81 (NR)	NR	0.57 (0.19, 0.95)
Hawkins-Kennedy	Apeldoorn et al. (2021)	0.68 (58.30, 76.20)	NR	0.33 (0.15, 0.51)
	Burns et al. (2016)	0.86 (NR)	NR	0.71 (0.41, 1.00)
	Ferenczi et al. (2018)	0.79 (NR)	NR	0.54 (0.22, 0.80)
	Lange et al. (2017)	0.75 (NR)	0.51 (NR)	0.47 (0.28, 0.67)
	Schmidt et al. (2021)	0.72 (NR)	NR	0.25 (0.01, 0.48)
Neer Walsh	Apeldoorn et al. (2021)	0.76 (66.90, 83.60)	NR	0.43 (0.23, 0.62)
	Burns et al. (2016)	0.75 (NR)	NR	0.51 (0.13, 0.88)
	Doxey et al. (2018)	0.90 (NR)	NR	NR
	Ferenczi et al. (2018)	0.88 (NR)	NR	0.64 (0.19, 0.91)
	Lange et al. (2017)	0.72 (NR)	0.56 (NR)	0.51 (0.33, 0.74)
Empty can	Apeldoorn et al. (2021)	0.76 (67.00, 83.40)	NR	0.51 (0.34, 0.66)
	Burns et al. (2016)	0.91 (NR)	NR	0.69 (0.28, 1.00)
	Doxey et al. (2018)	0.88 (NR)	NR	NR
	Lange et al. (2017)	0.86 (NR)	0.72 (NR)	0.69 (0.45, 0.97)
Full can	Apeldoorn et al. (2021)	0.83 (74.50, 89.20)	NR	0.62 (0.46, 0.87)
	Burns et al. (2016)	0.62 (NR)	NR	0.23 (0.00, 0.65)
Painful arc	Doxey et al. (2018)	0.93 (NR)	NR	NR

Test	Author	Agreement (95% CI)	PABAK (95% CI)	Kappa (95% CI)
Painful arc	Lange et al. (2017)	0.83 (NR)	0.66 (NR)	0.56 (0.37, 0.78)
	Schmidt et al. (2021)	0.70 (NR)	NR	0.56 (0.36, 0.76)
External rotation weakness	Apeldoorn et al. (2021)	0.75 (66.10, 82.70)	NR	0.50 (0.34, 0.66)
	Burns et al. (2016)	0.64 (NR)	NR	0.22 (0.00, 0.67)
	Doxey et al. (2018)	0.93 (NR)	NR	NR
	Schmidt et al. (2021)	0.88 (NR)	NR	0.70 (0.48, 0.92)
Internal rotation weakness	Apeldoorn et al. (2021)	0.77 (0.68, 0.84)	NR	0.47 (0.30, 0.65)
	Burns et al. (2016)	0.86 (NR)	NR	0.32 (0.00, 1.00)
ACJ palpation	Apeldoorn et al. (2021)	0.76 (67.00, 83.40)	NR	0.47 (0.29, 0.64)
Yergasson	Doxey et al. (2018)	0.95 (NR)	NR	NR
Belly press	Schmidt et al. (2021)	0.93 (NR)	NR	0.63 (0.25, 1.01)
External rotation lag sign	Schmidt et al. (2021)	0.89 (NR)	NR	0.40 (0.07, 0.74)
Internal rotation lag sign	Schmidt et al. (2021)	1.00 (NR)	NR	1.00 (1.00, 1.00)

Note: CI: confidence interval. ACJ: acromioclavicular joint. NR: Not recorded.

3.3.3 Diagnostic accuracy

The characteristics of the studies included in the diagnostic accuracy literature review are included in Table 6 below.

Table 6

Diagnostic accuracy study characteristics

Article	Design	No. of participants	Mean age	% Females	Clinical tests	Reference test
Anauate Nicolao et al., 2022	Prospective multicentre diagnostic accuracy study	199	47	42	Painful arc, full can, empty can, Hawkins Kennedy, resisted ER, Neer Walsh, drop arm	Arthroscopy
Ackmann et al., 2021	Prospective single-centre diagnostic accuracy study	61	60	43	Whipple test, empty can, full can	Arthroscopy
Balevi Batur et al., 2022	Prospective diagnostic accuracy study	106	55	68	Empty can, Neer Walsh, drop arm, Hawkins Kennedy, full can	MRI
Yazigi et al., 2021	Prospective, multicentre diagnostic accuracy study	575	51	48	Empty-can, full can, drop-arm test, Hawkins Kennedy, Neer Walsh, painful arc, resisted ER	MRI
Lädermann, Collin, et al., 2021	Diagnostic accuracy systematic review and meta-analysis	13 studies, 1699 participants	54	NR	Bear hug, belly press, IR lag sign, lift off test	Arthroscopic, MRI/MRA and ultrasound
Lädermann, Meynard, et al., 2021	Diagnostic accuracy systematic review and meta-analysis	14 studies, 2457 participants	54	NR	Painful arc, empty can, drop arm, Neer Walsh, Hawkins Kennedy, resisted ER, ER lag sign	Arthroscopic, MRI/MRA and ultrasound
Gismervik et al., 2017	Diagnostic accuracy systematic review and meta-analysis	11 articles, 4,125 participants	47	32	Neer Walsh, Hawkins Kennedy	Arthroscopic, MRI/MRA, CT, anaesthetic injection and ultrasound
Krill et al., 2018	Diagnostic accuracy systematic review	2 studies, 191 participants	NR	NR	Paxinos, ACJ palpation, active compression test, cross body adduction and Hawkins Kennedy	Anaesthetic injection

Note: No.: number. ER: external rotation. IR: internal rotation. MRI: magnetic resonance imaging. MRA: Magnetic resonance arthrography. CT: computed tomography scan. ACJ: acromioclavicular joint.

Subacromial pain diagnostic accuracy

Systematic reviews

There were 65 articles and 17 systematic reviews that matched the inclusion criteria for the diagnostic accuracy of subacromial pain. Of the 17 systematic reviews, three were selected that represented the most current research. The most recent systematic reviews were published in 2021 (Lädermann, Collin, et al., 2021; Lädermann, Meynard, et al., 2021). The systematic review by Lädermann, Meynard, et al. (2021) however, only included tests for posterior superior rotator cuff tears and Lädermann, Collin, et al. (2021) only addressed subscapularis tears. Gismervik et al. (2017) conducted the next most recent review, covering all types of subacromial pain and was therefore included with the two above reviews. All three of the selected reviews demonstrated 23 or higher out of 27 on the PRISMA scale with sound methodology (see above in Table 4).

Original articles

The Lädermann, Collin, et al. (2021) and Lädermann, Meynard, et al. (2021) reviews conducted their literature searches in 2020, therefore to update the systematic reviews, original articles that were published in 2021 or later were included. This resulted in four articles that matched the inclusion criteria (Ackmann et al., 2021; Anauate Nicolao et al., 2022; Balevi Batur et al., 2022; Yazigi et al., 2021). All four original articles had a low risk of bias on the QUADAS tool (see Table 7), they were well written articles that described the methods around their reference and index tests clearly.

Table 7

QUADAS ratings of diagnostic accuracy studies

QUADAS	Anauate Nicolao et al. (2022)	Ackmann et al., 2021	Balevi Batur et al., 2022	Yazigi et al., 2021
Patients were representative of standard practice	X	✓	✓	✓
Clear selection criteria	✓	✓	✓	✓
Appropriateness of reference test	✓	✓	✓	✓
Appropriate time between index and reference tests	✓	✓	?	✓
Reference test applied to the whole sample	✓	✓	✓	✓
The same reference test regardless of index test results	✓	✓	✓	✓
Reference test independent of the index test	✓	✓	✓	✓
Index test described in sufficient detail	✓	✓	✓	✓
Reference test described in sufficient detail	✓	X	✓	✓
The index test results were interpreted without knowledge of the reference test results	✓	✓	?	✓
The reference test results were interpreted without knowledge of the index test results	X	?	✓	✓
The same clinical data was available when test results were interpreted as would be available when the test is used in practice	✓	✓	✓	✓
Uninterpretable/ intermediate test results were reported	✓	X	?	✓
Withdrawals explained	✓	X	✓	✓
Total:	12/14	10/14	11/14	11/14

Note: ✓: positive finding. X: negative finding. ?: unclear finding.

Systematic review outcomes

Lädermann, Collin, et al. (2021) identified four tests that were eligible to perform a meta-analysis. These were the belly press, bear hug, internal rotation lag sign and the lift off test. They set a threshold of >0.80 specificity and sensitivity for a test to reach significance. No tests analysed reached this threshold however, and they concluded that no stand-alone test was able to accurately diagnose a subscapularis tear. Lädermann, Meynard, et al. (2021) in their review were able to perform a meta-analysis on the six following tests. These were the

Jobe test, Hawkins Kennedy, drop arm sign, painful arc, Neer Walsh test and the external rotation lag sign. Lädermann, Meynard, et al. (2021) have reached a similar overall conclusion to the above review. They summarized that no single shoulder test was able to adequately diagnose a specific tear of the supraspinatus, infraspinatus and/or teres minor. When testing the rotator cuff muscles as a group rather than individually however, tests showed improved diagnostic accuracy. The drop arm and external rotation lag sign showed specificities of 0.97 (CI 95% 0.62, 1.00) and 0.91 (CI 95% 0.50, 0.99) respectively for identifying rotator cuff tears. The pooled sensitivity of the empty can test was 0.77 (CI 95% 0.67, 0.85). There were no recorded likelihood ratios however to show the true size of the shift in probability.

The systematic review by Gismervik et al. (2017) also reached the same outcome; with no likelihood ratios reaching statistical significance. They concluded that no tests demonstrated adequate diagnostic accuracy and therefore the performance of stand-alone clinical tests for the diagnosis of subacromial pain is limited.

Original article outcomes

The original articles comparatively had more variable outcomes; these are outlined below in Table 8. The study by Anauate Nicolao et al. (2022) demonstrated significantly high positive likelihood ratios (+LR) >6 for identifying supraspinatus tears with empty and full can, drop arm and resisted external rotation tests. The only other study that reported significant likelihood ratios was Yazigi et al. (2021) with a +LR of 7.41 for the drop arm test. The other two original studies did not produce any +LRs >5, which is the threshold for a moderate shift in probability (McGee, 2002). Anauate Nicolao et al. (2022) also presented the only negative likelihood ratio that represented a moderate shift in probability (≤0.20). This was a painful arc demonstrating a sensitivity of 0.85 (CI 95% 0.79, 0.90) and a -LR of 0.20. Anauate Nicolao et al. (2022) however only included participants who had an indication for arthroscopic surgery, and may have already had a pathological diagnosis, leading to selection bias. Furthermore, neither Anauate Nicolao et al. (2022) or Yazigi et al. (2021) reported confidence intervals for their likelihood ratios.

Table 8

Subacromial diagnostic accuracy articles

Test	Author	Sp (95% CI)	Sn (95% CI)	+LR (95% CI)	-LR (95% CI)	DOR (95% CI)	PPV (95% CI)	NPV (95% CI)
Painful arc	Anauate Nicolao et al. (2022)	0.73 (0.60, 0.84)	0.85 (0.79, 0.90)	3.22 (NR)	0.20 (NR)	16.27 (NR)	0.91 (NR)	0.60 (NR)
	Lädermann, Meynard, et al. (2021)	0.74 (0.46, 0.90)	0.51 (0.29, 0.72)	NR	NR	NR	0.60 (0.49, 0.69)	0.66 (0.58, 0.74)
	Yazigi et al. (2021)	0.38 (NR)	0.78 (NR)	1.26 (NR)	0.57 (NR)	2.20 (NR)	0.65 (NR)	0.54 (NR)
Empty can/Jobe test	Anauate Nicolao et al. (2022)	0.93 (0.81, 0.98)	0.75 (0.63, 0.84)	10.05 (NR)	0.27 (NR)	40 (NR)	0.94 (NR)	0.70 (NR)
	Ackmann et al. (2021)	0.65 (NR)	0.77(NR)	NR	NR	NR	0.85 (NR)	0.48 (NR)
	Balevi Batur et al. (2022)	0.34 (0.22, 0.48)	0.76 (0.63, 0.86)	1.16 (0.91, 1.46)	0.70 (0.39, 1.35)	NR	0.54 (0.48, 0.59)	0.59 (0.44, 0.72)
	Lädermann, Meynard, et al. (2021)	0.67 (0.59, 0.73)	0.77 (0.67, 0.85)	NR	NR	NR	0.72 (0.54, 0.85)	0.73 (0.58, 0.85)
	Yazigi et al. (2021)	0.81 (NR)	0.38 (NR)	1.30 (NR)	0.49 (NR)	2.64 (NR)	0.66 (NR)	0.58 (NR)
Full can	Anauate Nicolao et al. (2022)	0.91 (0.78, 0.96)	0.63 (0.50, 0.74)	6.76 (NR)	0.41 (NR)	16.53 (NR)	0.91 (NR)	0.63 (NR)
	Ackmann et al. (2021)	0.61 (NR)	0.66 (NR)	NR	NR	NR	0.78 (NR)	0.38 (NR)
	Balevi Batur et al. (2022)	0.31 (0.20, 0.45)	0.76 (0.63, 0.86)	1.10 (0.87, 1.37)	0.78 (0.43, 1.41)	NR	0.52 (0.47, 0.58)	0.56 (0.41, 0.70)
	Yazigi et al. (2021)	0.45 (NR)	0.75 (NR)	1.36 (NR)	0.55 (NR)	2.46 (NR)	0.67 (NR)	0.55 (NR)
Hornblower's test	Yazigi et al. (2021)	0.75 (NR)	0.44 (NR)	1.74 (NR)	0.75 (NR)	2.31 (NR)	0.72 (NR)	0.47 (NR)
Drop arm	Anauate Nicolao et al. (2022)	0.98 (0.89, 1.00)	0.19 (0.13, 0.25)	9.21 (NR)	0.83 (NR)	11.10 (NR)	0.97 (NR)	0.26 (NR)

Test	Author	Sp (95% CI)	Sn (95% CI)	+LR (95% CI)	-LR (95% CI)	DOR (95% CI)	PPV (95% CI)	NPV (95% CI)
Drop arm	Balevi Batur et al. (2022)	0.97(0.88, 1.00)	0.12 (0.05, 0.23)	3.50 (0.75, 16.14)	0.91 (0.82, 1.01)	NR	0.78 (0.43, 0.94)	0.52 (0.50, 0.55)
	Lädermann, Meynard, et al. (2021)	0.97 (0.62, 1.00)	0.38 (0.01, 0.98)	NR	NR	NR	0.71 (0.35, 0.92)	0.87 (0.29, 0.73)
	Yazigi et al. (2021)	0.98 (NR)	0.16 (NR)	7.41 (NR)	0.86 (NR)	8.63 (NR)	0.92 (NR)	0.44 (NR)
Neer Walsh	Anauate Nicolao et al. (2022)	0.82 (0.69, 0.90)	0.78 (0.71, 0.84)	4.26 (NR)	0.27 (NR)	15.92 (NR)	0.93 (NR)	0.53 (NR)
	Balevi Batur et al. (2022)	0.21 (0.11, 0.33)	0.69 (0.55, 0.80)	0.87 (0.70, 1.08)	1.50 (0.80, 2.83)	NR	0.47 (0.41, 0.52)	0.40 (0.26, 0.56)
	Gismervik et al. (2017)	0.60 (NR)	0.59 (NR)	1.48 (NR)	0.68 (NR)	2.17 (NR)	NR	NR
	Lädermann, Meynard, et al. (2021)	0.39 (0.06, 0.86)	0.70 (0.57, 0.81)	NR	NR	NR	0.63 (0.46, 0.77)	0.50 (0.14, 0.86)
	Yazigi et al. (2021)	0.58 (NR)	0.63 (NR)	1.49 (NR)	0.64 (NR)	2.34 (NR)	0.69 (NR)	0.51 (NR)
Hawkins Kennedy	Anauate Nicolao et al. (2022)	0.65 (0.51, 0.77)	0.80 (0.73, 0.85)	2.30 (NR)	0.31 (NR)	7.35 (NR)	0.88 (NR)	0.49 (NR)
	Balevi Batur et al. (2022)	0.22 (0.13, 0.35)	0.90 (0.79, 0.96)	1.16 (0.98, 1.36)	0.46 (0.19, 1.13)	NR	0.54 (0.50, 0.58)	0.68 (0.47, 0.84)
	Gismervik et al. (2017)	0.67 (NR)	0.58 (NR)	1.76 (NR)	0.63 (NR)	2.86 (NR)	NR	NR
	Lädermann, Meynard, et al. (2021)	0.49 (0.17, 0.81)	0.73 (0.50, 0.88)	NR	NR	NR	0.64 (0.49, 0.77)	0.59 (0.31, 0.82)
	Yazigi et al. (2021)	0.55 (NR)	0.65 (NR)	1.45 (NR)	0.64 (NR)	2.28 (NR)	0.68 (NR)	0.51 (NR)
Resisted ER	Anauate Nicolao et al. (2022)	0.95 (0.85, 0.99)	0.40 (0.31, 0.50)	8.89 (NR)	0.62 (NR)	14.25 (NR)	0.95 (NR)	0.43 (NR)
	Lädermann, Meynard, et al. (2021)	0.52 (NR)	0.69 (NR)	NR	NR	4.66 (NR)	0.31 (NR)	0.85 (NR)
	Yazigi et al. (2021)	0.58 (NR)	0.69 (NR)	1.61 (NR)	0.54 (NR)	2.96 (NR)	0.71 (NR)	0.55 (NR)

Test	Author	Sp (95% CI)	Sn (95% CI)	+LR (95% CI)	-LR (95% CI)	DOR (95% CI)	PPV (95% CI)	NPV (95% CI)
Whipple test	Ackmann et al. (2021)	0.41 (NR)	0.80 (NR)	NR	NR	NR	0.78 (NR)	0.44 (NR)
Bear hug	Lädermann, Collin, et al. (2021)	0.94 (0.80, 0.99)	0.55 (0.28, 0.79)	NR	NR	NR	0.82 (0.63, 0.93)	0.80 (0.68, 0.89)
Belly press	Lädermann, Collin, et al. (2021)	0.94 (0.77, 0.99)	0.48 (0.29, 0.68)	NR	NR	NR	0.76 (0.50, 0.90)	0.80 (0.70, 0.87)
IR lag sign	Lädermann, Collin, et al. (2021)	0.92 (0.73, 0.98)	0.32 (0.13, 0.61)	NR	NR	NR	0.58 (0.31, 0.82)	0.75 (0.62, 0.85)
Lift off	Lädermann, Collin, et al. (2021)	0.94 (0.81, 0.98)	0.33 (0.20, 0.51)	NR	NR	NR	0.70 (0.44, 0.87)	0.76 (0.68, 0.82)
ER lag sign	Lädermann, Meynard, et al. (2021)	0.91 (0.50, 0.99)	0.46 (0.03, 0.96)	NR	NR	NR	0.47 (0.20, 0.76)	0.88 (0.58, 0.97)

Note: Sp: specificity. Sn: sensitivity. +LR: positive likelihood ratio. -LR: negative likelihood ratio. DOR: diagnostic odds ratio. PPV: positive predictive value.

NPV: negative predictive value. CI: confidence interval. ACJ: acromioclavicular joint. NR: Not recorded. ER: external rotation. IR: internal rotation

With the exception of the above five tests, the remaining OSTs included in these original studies demonstrated insufficient diagnostic accuracy to identify subacromial pain based on the likelihood ratios. It is important to consider the likelihood ratios when interpreting results as they are a more credible measure of diagnostic accuracy than specificity and sensitivity (Denegar & Fraser, 2006). Positive LRs of <5 and -LRs of >0.20 only demonstrate a small shift in probability of a condition being present or absent (Denegar & Fraser, 2006). This level of change in probability is not sufficient to consider the test findings meaningful in diagnosing subacromial pain in a clinical setting.

Acromioclavicular joint pain diagnostic accuracy

Systematic review

Six original studies on the diagnostic accuracy of CACJ pain were identified; however no new studies have been published since 2013. The systematic review by Krill et al. (2018) therefore represents the most current summary of evidence. As it is of moderate quality on the PRISMA score (see Table 4 above) and reviews the same articles found our search, no further review was required. Krill et al. (2018) scored 19/27 on the PRISMA scale, indicating moderate quality. The main PRISMA criteria this review did not meet were the clarity of reporting biases, and outlining the methods used to assess bias. Krill et al. (2018) only included two original articles in their review, as the remaining studies found did not reach the criteria of level I or II evidence. Only Cadogan, McNair, Laslett and Hing (2013) and Walton et al. (2004) demonstrated a sound standard of reporting and control of bias. Table 9 and 10 below outline the findings of the two included articles in the Krill et al. (2018) systematic review.

Table 9

Individual ACJ tests from Krill et al., 2018

Test	Author	Sp (95% CI)	Sn (95% CI)	+LR (95% CI)	-LR (95% CI)	PPV (95% CI)	NPV (95% CI)
Active compression test (O'Brien's test)	Cadogan, McNair, Laslett and Hing (2013)	0.92 (0.19, 0.35)	0.14 (0.05, 0.33)	1.73 (0.58, 1.12)	0.94 (0.71, 2.43)	0.23 (0.08, 0.21)	0.86 (0.66, 0.90)
	Walton et al. (2004)	0.90 (NR)	0.16 (NR)	1.60 (NR)	0.93 (NR)	0.62 (NR)	0.52 (NR)
Paxinos test	Walton et al. (2004)	0.50 (NR)	0.79 (NR)	1.58 (NR)	0.42 (NR)	0.61 (NR)	0.70 (NR)
Acromioclavicular joint tenderness	Cadogan, McNair, Laslett and Hing (2013)	0.73 (0.65, 0.80)	0.36 (0.20, 0.57)	1.37 (0.70, 2.39)	0.87 (0.58, 1.13)	0.19 (0.10, 0.33)	0.87 (0.79, 0.92)
	Walton et al. (2004)	0.10 (NR)	0.96 (NR)	1.07 (NR)	0.40 (NR)	0.52 (NR)	0.71 (NR)
Cross body adduction	Cadogan, McNair, Laslett and Hing (2013)	0.26 (0.19, 0.35)	0.64 (0.43, 0.80)	1.39 (0.58, 1.12)	0.86 (0.71, 2.43)	0.13 (0.08, 0.21)	0.81 (0.66, 0.90)
Pain present below elbow	Cadogan, McNair, Laslett and Hing (2013)	0.18 (0.12, 0.26)	1.00 (0.84, 1.00)	1.22 (1.18, 1.34)	0.00 (0.00, 0.92)	0.16 (0.11, 0.23)	1.00 (0.86, 1.00)
Hawkins-Kennedy	Cadogan, McNair, Laslett and Hing (2013)	0.36 (0.28, 0.44)	0.70 (0.48, 0.86)	1.09 (0.74, 1.41)	0.84 (0.39, 1.55)	0.15 (0.09, 0.23)	0.88 (0.77, 0.95)

Note: Sp: specificity. Sn: sensitivity. +LR: positive likelihood ratio. -LR: negative likelihood ratio. PPV: positive predictive value NPV: negative predictive value.

NR: not recorded.

Table 10

Combined acromioclavicular joint tests from Krill et al., 2018

Author	Tests	Variables	Sp (95% CI)	Sn (95% CI)	+LR (95% CI)	-LR (95% CI)	PPV (95% CI)	NPV (95% CI)
Cadogan, McNair, Laslett and Hing (2013)	• Pain onset due to repetitive activity • No pain below elbow • ACJ thickened or swollen • No pain end-range passive GHJ abduction • No pain end range passive ER (90° abduction)	1/5	0.07 (0.04, 0.13)	1.00 (0.85, 1.00)	1.08 (1.07, 1.15)	0.00 (0.00, 2.15)	0.16 (0.11, 0.23)	1.00 (0.70, 1.00)
		4/5	0.95 (0.90, 0.98)	0.23 (0.10, 0.43)	4.98 (1.69, 13.84)	0.81 (0.59, 0.95)	0.46 (0.21, 0.72)	0.88 (0.82, 0.92)
Cadogan, McNair, Laslett and Hing (2013)	• Cross-body adduction • Active compression/ • O'Brien's test • Hawkins- Kennedy test • Localised ACJ tenderness (palpation)	1/4	0.07 (0.04, 0.13)	0.96 (0.78, 0.99)	1.03 (0.84, 1.11)	0.65 (0.84, 1.11)	0.15 (0.10, 0.22)	0.90 (0.60, 0.98)
		4/4	0.99 (0.95, 1.00)	0.05 0.01, 0.24)	5.70 (0.60, 52.63)	0.96 (0.77, 1.01)	0.50 (0.10, 0.91)	0.86 (0.79, 0.91)
Walton et al. (2004)	• Cross body adduction stress test • AC resisted extension • Active compression test (palpation)	1/3	0.74 (NR)	0.00 (NR)	NR	NR	0.31 (NR)	0.96 (NR)
		3/3	0.97 (NR)	0.25 (NR)	NR	NR	0.17 (NR)	1.00 (NR)

Note: Sp: specificity. Sn: sensitivity. +LR: positive likelihood ratio. -LR: negative likelihood ratio. PPV: positive predictive value. NPV: negative predictive value. NR: not recorded. GHJ: glenohumeral joint. ER: external rotation. AC: acromioclavicular. ACJ: acromioclavicular joint.

Outcomes

The diagnostic accuracy of OSTs for CACJ pain reflect similar outcomes to the subacromial pain tests. Krill et al. (2018) concluded that no OSTs in any combination demonstrated more than small likelihood ratios for diagnosing or excluding an ACJ pathology. The strongest combination of OSTs were the Paxinos and O'Brien's tests that when performed sequentially in series demonstrated 0.96 specificity. When tested in parallel if one of the tests was positive then the sensitivity was 0.94. They also showed the highest +LR of 2.71 and -LR of 0.35. These likelihood ratios only represent a small shift in probability of 15-30% of ruling ACJ pain in or out. No confidence intervals or p-values were recorded.

Cadogan, McNair, Laslett and Hing (2013) demonstrated that a cluster of five variables from the patient interview, range of motion tests and clinical observations, were more useful than any combination of special tests. These were: pain onset from a repetitive mechanism, no pain below the elbow, an ACJ deformity on observation, no pain with passive glenohumeral (GHJ) abduction, and no pain with passive GHJ external rotation at 90 degrees abduction. An absence of four of the above variables had a -LR of 0.00 (CI 95% 0.00, 2.15) and a mean estimate sensitivity of 1.00 (CI 95% 0.85, 1.00). When four of the five variables were positive the specificity was 0.95 (CI 95% 0.90, 0.98) and the positive likelihood ratio was 4.98 (CI 95% 1.69, 13.84). The likelihood ratio confidence intervals cross one however, inferring these findings did not reach statistical significance. Pain present below the elbow was the only stand-alone measure that showed statistically significant outcomes (Cadogan, McNair, Laslett, & Hing, 2013). The mean estimate of sensitivity was 1.00 (CI 95% 0.84, 1.00), and the -LR was 0.00 (CI 95% 0.00, 0.92); therefore if participants had pain below the elbow, they could confidently rule out ACJ pain as a potential diagnosis in their cohort.

3.4 Discussion

The aim of this review was to explore and critique the existing literature on the validity of OSTs for the diagnosis of ACJ or subacromial pain.

Five systematic reviews (Gismervik et al., 2017; Krill et al., 2018; Lädermann, Collin, et al., 2021; Lädermann, Meynard, et al., 2021; Lange et al., 2017) and eight original studies (Ackmann et al., 2021; Anauate Nicolao et al., 2022; Apeldoorn et al., 2021; Balevi Batur et al., 2022; Burns et al., 2016; Ferenczi et al., 2018; Schmidt et al., 2021; Yazigi et al., 2021) met the inclusion criteria.

The majority of the included studies questioned the ability of OSTs to diagnose ACJ and subacromial pathologies. There are over 70 OSTs for the shoulder that have been designed in the attempt to differentiate and diagnose specific shoulder structures (Salamh & Lewis, 2020). One of the factors that make it difficult to identify an isolated pain source in the shoulder, is the complicated anatomy in this region (Clark et al., 1990; Clark & Harryman, 1992). The rotator cuff sits in close proximity to and interacts with the shoulder capsule, ACJ, ligaments and bursa (Clark & Harryman, 1992). Hence, a test that provokes pain in the shoulder is likely to be loading multiple structures. The following sections explore the clinical considerations of our review in relation to other published research.

3.4.1 Reliability of special tests

The first question this review addressed was if OSTs for the shoulder have sufficient reliability to be used in a clinical setting? Based on the reported findings of the studies included, the majority of OSTs have only fair to moderate reliability and should therefore be used with caution in a clinical setting, if at all. Only the belly press and internal rotation lag sign demonstrated substantial to almost perfect inter-rater reliability with a kappa between 0.63-1.00. The Neer Walsh and empty can tests produced moderate to substantial reliability (0.43-0.69 kappa). The reliability of the Hawkins Kennedy and resisted external rotation tests varied significantly from low kappa scores to high. The remaining tests examined demonstrated low to moderate agreement; with resisted external rotation, the full can test and Hawkins Kennedy demonstrating the lowest kappa values across the studies.

Variability in the research

The systematic reviews and meta-analysis by Lange et al. (2017) and May et al. (2010) both report large variances in the quality of previous studies and the inter-rater reliability outcomes. Possible reasons for varying reliability results may be time differences between the testing, differences in recruited cohorts, a participant learning affect, aggravated symptoms with multiple tests, differences in the experience of the raters and/or different interpretations of the performance of the tests (Cadogan, Laslett, et al., 2011a; Lange et al., 2017). There are also a large number of different OSTs and each study only reviews a small portion of these, therefore comparisons are not always possible (Salamh & Lewis, 2020).

The significant variability in outcomes is evident in our review. Three tests reviewed ranged from fair to substantial reliability across the studies. These were the Hawkins Kennedy test (0.25 to 0.71), the full can test (0.23 to 0.62), and external rotation weakness (0.22 to 0.70) (Landis & Koch, 1977). There are also variations in the criteria used to interpret kappa outcomes (McHugh, 2012). The Landis and Koch (1977) criteria rates 0.21-0.40 as fair and 0.41-0.60 as moderate; McHugh (2012) in contrast presents any kappa value <0.60 as weak and ≥0.60 as moderate. This makes reaching an overarching conclusion on the reliability of OSTs difficult.

Comparisons to current research

There are comparisons that can be drawn from the outcomes in our review to other published research. The painful arc test in our review demonstrated a kappa of 0.56; this finding is supported by previous reliability studies which have reported a moderate to substantial agreement for this test (Michener et al., 2009; Nomden et al., 2009; Ostor et al., 2004). This indicates it may be a reliable test to use in the diagnosis of ACJ or subacromial pain.

The reliability of the empty can test varied in our review with a kappa from 0.51-0.69. This is similar to the outcomes reported in the previous research which range from moderate to near perfect agreement (Holtby & Razmjou, 2004; Johansson & Ivarson, 2009; Michener et al., 2009; Ostor et al., 2004). Despite the variance seen across these studies, this test has

arguably shown adequate reliability to be utilised in a clinical setting based on the Landis and Koch (1977) criteria.

The wide variance seen in the outcomes of the Hawkins Kennedy test in our review (kappa 0.25 to 0.71) is similar to other study's findings. Michener et al. (2009) reported outcomes of fair agreement for Hawkins Kennedy test in direct contrast to Johansson and Ivarson (2009) who described a near perfect agreement. Based on these conflicting findings, this is not a consistent test and therefore it cannot be recommended for clinical use.

The larger question is whether research should continue to examine the reliability of OSTs or abandon them altogether? Our systematic review agrees with findings from a systematic review and meta-analysis by May et al. (2010). They summarised that the overall evidence for the reliability of tests was contradictory and that only half of the studies 36 studies included in their review were of high quality. Those studies of acceptable quality were less likely to demonstrate sufficient reliability scores. May et al. (2010) concluded that from the studies reviewed, the levels of reliability were insufficient to ensure reproducible diagnostic processes and therefore could not be recommended for use in clinical practice.

There is a lack of consistent evidence to show that OSTs have adequate reliability in a clinical setting (Cadogan, Laslett, et al., 2011a; Nurjannah & Siwi, 2017). Unreliable inter-examiner tests directly affect the interpretation of clinical findings and accordingly may confound the diagnostic process (Lange et al., 2017). A test loses its value if the findings cannot be relied on. A wide body of research supports the conclusion that the majority of OSTs are unreliable tools and their use in the clinical setting needs to be re-evaluated (Hegedus et al., 2017; Lange et al., 2017; May et al., 2010; Salamh & Lewis, 2020).

3.4.2 Diagnostic accuracy of subacromial special tests

The second question of this review addressed the diagnostic accuracy of subacromial tests. The systematic reviews by Gismervik et al. (2017) and Lädermann, Collin, et al. (2021) showed that no OSTs reached statistical significance for diagnosing subacromial pain or subscapularis tears. Lädermann, Meynard, et al. (2021) concluded that no OST had adequate diagnostic accuracy to diagnose specific tears of the posterior rotator cuff. When

grouping the rotator cuff tendon tears together as a diagnosis however, the drop arm and Neer Walsh tests demonstrated specificity over 0.90 and the empty can test a sensitivity of over 0.77. Both had adequate confidence intervals to demonstrate statistical significance; unfortunately, no likelihood ratios were recorded.

The original diagnostic accuracy articles were more varied in their conclusions. Two of the four studies demonstrated high levels of diagnostic accuracy for the following tests. The study by Anauate Nicolao et al. (2022) produced statistically significantly positive likelihood ratios (>6) for identifying supraspinatus tears with empty and full can, drop arm and resisted external rotation tests. Yazigi et al. (2021) demonstrated a +LR of 7.41 for the drop arm test. Additionally, Anauate Nicolao et al. (2022) produced a -LR of 0.20 for the painful arc test; this signifies a moderate shift in probability. Of note these two studies were performed by the same authors. Neither of these studies reported confidence intervals, and the 2022 study demonstrated a significant selection bias by only including participants allocated for rotator cuff surgery.

In comparison, the study by Balevi Batur et al. (2022) reported that no tests had likelihood ratios that reached statistical significance for the diagnosis of subacromial pain. Ackerman et al. (2018) produced a 0.77 sensitivity for the empty can test, and a 0.80 for the Whipple test but did not report LRs or CIs. These outcomes should be interpreted with caution; without the other statistical measures, the credibility of these results cannot be confidently established. Of note the tests that presented with higher measures of diagnostic utility were more commonly measures of strength.

This review has concluded that the inconsistent reporting of diagnostic measures, such as likelihood ratios and confidence intervals, in the included studies makes it difficult to reach a decision on the diagnostic accuracy of the OSTs. Without robust methodology and complete statistical reporting, the small handful of tests that produced significant results cannot be taken at face value. Overall, the results point to no individual orthopaedic special tests demonstrating adequate diagnostic accuracy to differentiate subacromial pain or specific rotator cuff tears from other structures in the shoulder.

This outcome mirrors the results of two well-designed systematic reviews by Hegedus et al. (2008) and Green et al. (2008). Later reviews by Hegedus et al. (2015) and Hegedus et al. (2017) have also summarized that despite the large body of research on OSTs, many of the articles are of insufficient quality. Of those that are high quality, very few demonstrate adequate statistical significance for OSTs to be valid tools in the clinical setting.

Cadogan et al. (2016) concluded in their diagnostic accuracy study that few clinical variables lead to the accurate diagnosis of subacromial pain. The three tests that when combined showed the highest and lowest likelihood ratios were anterior shoulder pain, a strain mechanism of injury, and an absence of pain in end range abduction/external rotation. This adds to the growing body of evidence that the patient interview, range of motion, and strength tests may be more accurate and reliable to diagnose a patient's shoulder complaint than OSTs (Hegedus et al., 2015; Hegedus et al., 2017; Raynor & Kuhn, 2016; Salamh & Lewis, 2020; Schmidt et al., 2021).

3.4.3 Diagnostic accuracy of acromioclavicular special tests

The third section of this review reported on the diagnostic accuracy of OSTs for an ACJ pathology. The evidence for diagnosing CACJ pathology is limited, especially in comparison to subacromial pain. Six relevant original articles were found in this review on the diagnostic accuracy of ACJ pain compared to 65 relevant studies addressing subacromial pain. This highlights the disparities which carry into the clinical setting where subacromial pain accounts for approximately 44-65% of presenting shoulder complaints (Consigliere et al., 2018). As a result, CACJ pain is frequently overlooked and misdiagnosed in primary care (Cadogan, McNair, Laslett, & Hing, 2013; Javed et al., 2017).

The most recent systematic review by Krill et al. (2018) concluded that no OST was accurate in the diagnosis of CACJ pain. The strongest combination of OSTs presented by Krill et al. (2018) were the Paxinos and O'Brien's tests which had the highest +LR of 2.71 and -LR of 0.35. These likelihood ratios only represent a small shift in probability. One of the included studies by Cadogan, McNair, Laslett and Hing (2013) did show a -LR of 0.00 (95% CI 0.00, 2.15) and a +LR of 4.00 (95% CI 1.69, 13.84) with their cluster of five variables. Of note, the confidence intervals were wide and crossed one, indicating uncertain estimates and no statistical significance.

Biederwolf (2013) performed a systematic review to identify clusters of OSTs for the diagnosis of ACJ pain. Similar to our review, they concluded that stand alone tests demonstrated insufficient diagnostic accuracy to identify an ACJ pathology. They proposed a cluster of tests instead, including the cross-body adduction test, resisted AC extension and the O'Brien's test for the diagnosis of ACJ pain. This had a post-test probability of 80.5% when all tests were positive. Powell and Huijbregts (2006) recommended the same three tests in their review to achieve a +LR of 8.30, based on the same original articles. A cluster of three tests (when negative) to rule out ACJ pain was also proposed by Powell and Huijbregts (2006), consisting of the cross-body adduction test, tenderness on palpation of the ACJ, and the Paxinos sign. Despite our review not demonstrating any statistically significant likelihood ratios for the above tests individually, it is possible when combined these tests have improved diagnostic accuracy (Hegedus et al., 2015). Clusters of tests such as these ones would require further research to validate their merit for the diagnosis of ACJ pain.

Studies which display high likelihood ratios for stand-alone OSTs need to be examined carefully for quality and level of bias. Two original studies examining the diagnostic accuracy of ACJ tests, O'Brien et al. (1998) and Chronopoulos et al. (2004), demonstrated a +LR of 13.30 and 8.20 respectively for the active compression test (O'Brien's test). These studies however did not meet the threshold to be included in the Krill et al. (2018) systematic review as they were not level I or II evidence. The two studies that were included, Cadogan, McNair, Laslett and Hing (2013) and Walton et al. (2004), demonstrated +LRs of 1.73 and 1.60 respectively for the active compression test. These lower likelihood ratios may therefore be the more credible outcomes to base clinical decisions on.

Hegedus et al. (2008) in their systematic review reached the same conclusion as our review; OSTs have inadequate diagnostic accuracy to be used as individual tests. They rationalised that OSTs are not able to separate the ACJ from other shoulder structures due to close proximity and that an ACJ pathology can also be part of a combined diagnosis leading to confusion when testing.

3.4.4 Limitations of the included studies

The identified limitations below put the results into context and outline the quality and biases of the studies included (Ioannidis, 2007).

Reliability studies

The assessment of reliability using kappa values does not adjust for differences in prevalence and bias in the testing population. A commonly used measure to overcome this is the PABAK score (Byrt et al., 1993; Chen et al., 2009). Of the five articles reviewed only Lange et al. (2017) reported PABAK values. Unfortunately, they did not report confidence intervals to determine statistical significance of the estimates. Additionally, two of the reliability studies did not clearly report their blinding processes and whether they randomised the order of their testing (Ferenczi et al., 2018; Schmidt et al., 2021). This introduces the possibility of performance and detection bias.

Subacromial diagnostic accuracy studies

The studies included for the diagnostic accuracy of subacromial pain each used a reference test of either an MRI and/or arthroscopic visualization. The one exception was the systematic review by Gismervik et al. (2017) which included studies utilising subacromial anaesthetic injections. Diagnostic injections are the accepted gold standard reference test to identify a specific source of shoulder pain (Cardone & Tallia, 2002; McFarland et al., 2017). An MRI can detect just as many abnormalities in asymptomatic shoulders as they can in those with pain, and surgical findings do not always detect the primary pain generator (Barreto et al., 2019; Hegedus et al., 2017; Lewis et al., 2015). If a reference test is flawed then any comparison made against it may also be unreliable (Hegedus et al., 2017).

Ackmann et al. (2021) and Anauate Nicolao et al. (2022) only accepted participants who were presenting for arthroscopic surgery, and in both studies by Balevi Batur et al. (2022) and Ackmann et al. (2021) only participants who had confirmed rotator cuff tears on MRI were included. This is selection bias, and it limits the generalisability of results and the reported sensitivity and likelihood ratios of their outcomes.

The systematic reviews by Lädermann, Collin, et al. (2021) and Lädermann, Meynard, et al. (2021) did not record likelihood ratios. These are important measures that help the reader determine the shift in probability of a condition being present or absent. Anauate Nicolao et al. (2022) only recorded confidence intervals for specificity and sensitivity but not for other values. Confidence intervals are necessary when interpreting research findings as they help to explain the precision and statistical significance of the outcomes.

Acromioclavicular joint diagnostic accuracy studies

When presenting the results for the diagnostic accuracy of ACJ tests, Krill et al. (2018) did not record any confidence intervals or p-values. Additionally, the two ACJ studies assessed by Krill et al. (2018) had significant variations in their PAR thresholds affecting comparability. The PAR criteria used by Cadogan, McNair, Laslett and Hing (2013) was an 80% or higher reduction in pain compared to the Walton et al. (2004) study which set a 50% or higher PAR threshold. Both studies had low numbers reaching a PAR, with 22 and 28 participants respectively, limiting their statistical power.

3.4.5 Strengths of this review

To the researcher's knowledge this is the only systematic review that addresses both the reliability and diagnostic accuracy of shoulder tests. This review fills a gap in the literature and summarizes the current knowledge on the validity of commonly utilised shoulder assessment tools.

3.4.6 Limitations of this review

The limitations of this review are that only studies published in English were sourced and one researcher conducted the search strategy and screened the titles and abstracts. It is possible there are studies that have been overlooked or excluded due to bias. Studies published prior to 1990 were excluded due to pragmatic reasons; the objective was to update current systematic reviews and earlier studies were not required for this. Restricting the dates in the search strategy does however increase the possibility of a systematic bias (Helbach et al., 2022).

3.4.7 Clinical implications

This review corresponds with many other systematic reviews which conclude that the reliability and diagnostic accuracy of OSTs are poor. OSTs are unable to identify specific symptomatic structures in the shoulder. An inaccurate diagnosis can result in ineffective treatment, which can include interventions with associated risks, such as steroid injections and/or surgical management (Salamh & Lewis, 2020). A thorough patient interview and the use of strength and range of motion tests may therefore be a superior method of reaching a diagnosis in clinical practice (Salamh & Lewis, 2020). Combining variables from the wider

clinical assessment has been shown to be more accurate than the use of stand-alone measures (Hegedus et al., 2015). A diagnostic anaesthetic injection may add further certainty to assist in clinical decision making (Cadogan et al., 2012).

3.4.8 Recommendations for future research

The reliability and diagnostic accuracy of most OSTs for the diagnosis of ACJ and subacromial pain are inadequate and cannot be recommended for use in clinical practice. Further research on the validity of OSTs is not warranted. Given the findings from this review the proposal for future research would be to shift the focus to the validity of clusters of clinical tests other than OSTs. This more closely resembles how they are utilised in clinical practice (Hegedus et al., 2015; May et al., 2010). Additionally, it would be recommended to focus less on pain provoking tests in research and shift towards more objective tools such as range of motion and strength tests (Lange et al., 2017; Lewis, 2009).

3.5 Conclusion

It is no longer acceptable practice to rely on the outcomes of OSTs alone to reach a clinical diagnosis. The reliability of OSTs is variable with poor to moderate agreement at best. The exception to this rule, with high levels of reliability, was the belly press and internal rotation lag sign tests. These are both measures of strength which have been demonstrated to be more reliable than reported pain scores.

No individual OSTs have sufficient accuracy to identify or exclude a subacromial or CACJ pathology as the primary source of pain. This review adds strength to the argument that OSTs have poor validity and are not able to distinguish between pain sensitive structures in the shoulder. Clinicians should place less trust in these special tests when making clinical decisions concerning patient management. Evidence from our review and current research suggests that more weight should be given to the patient interview questions, along with range of motion and strength tests rather than OSTs. The flaws highlighted in these studies bring to light the necessity of further research being conducted with robust methodology focussing on the validity of tools other than OSTs to diagnose shoulder complaints. Diagnostic accuracy studies should focus on the combination of these tools rather than stand-alone tests.

Chapter 4 Reliability of Orthopaedic Special Tests for Acromioclavicular Joint and Subacromial Pain.

4.1 Introduction

The shoulder consists of four joints and multiple soft tissue structures that are capable of generating pain (Chang et al., 2022). Accurately diagnosing the primary pain source in the shoulder can enable a clinician to tailor the patient's treatment plan (Cadogan, McNair, Laslett, & Hing, 2013). For example determining if a cortisone injection would be appropriate and where to place it, whether a patient would benefit from a strengthening programme or whether a surgical opinion is warranted (ACC, 2004).

A misdiagnosis can lead to chronic pain and have a negative effect on a patient's mood, sleep, quality of life and overall health status (Fine, 2011). Additionally an incorrect diagnosis can result in excessive imaging, inappropriate interventions, time off work and further visits to medical practitioners (Deyo, 2002). These adverse outcomes highlight the importance of the assessment process in reaching an accurate conclusion (Deyo, 2002).

Currently, a clinical diagnosis is made on the basis of information obtained from the patient interview and physical examination, plus or minus medical imaging (Shaffer, 1999). This however relies on the clinical tests being valid to accurately measure what the clinician is intending to assess. The current reliability and diagnostic accuracy research is variable, indeterminate and questions the usefulness of individual tests in their ability to identify specific shoulder pathologies (Gismervik et al., 2017; Hegedus et al., 2008; Lange et al., 2017).

To be valid a clinical test requires adequate diagnostic accuracy and reliability (Fritz & Wainner, 2001; Šimundić, 2009). There are two types of reliability measures; intra-rater reliability looks at the consistency of the outcomes collected by one examiner over more than one attempt of a test and inter-rater reliability measures the variability of results between two or more examiners assessing the same test (Scheel et al., 2018).

Poor inter-rater agreement suggests that the findings of a test should not be relied on as the results vary from rater to rater, this can lead to different conclusions about the presence or absence of the condition of interest (Apeldoorn et al., 2021). Poor reliability therefore can diminish the value of an individual or cluster of tests, as a test is not valid if it cannot be consistent in its' measurement (Lange et al., 2017; May et al., 2010).

While diagnostic accuracy studies for orthopaedic special tests (OST) of the shoulder are common, there is limited research on the reliability of these tests, especially for the ACJ (Cadogan, Laslett, et al., 2011a; Lange et al., 2017). The reliability of many diagnostic tests for the shoulder have not been investigated (Lange et al., 2017). The tests that have been presented in the literature show variable reliability with the majority of them demonstrating low to moderate inter-rater reliability (Lange et al., 2017; May et al., 2010). Our systematic review found similar findings, emphasising the need for clinicians to be more cautious when basing clinical decisions on the outcomes of these individual tests.

Hence, this study investigated the inter-rater reliability of commonly utilised orthopaedic special tests that have previously not been examined for ACJ and subacromial pain. This study's purpose was to determine the reliability of these tests to inform the decision about whether to include them in the subsequent diagnostic accuracy study.

4.2 Methodology

4.2.1 Study design

This was a prospective test-retest inter-examiner reliability study of clinical tests for the shoulder. This study and the diagnostic accuracy study were registered in advance with the Australian New Zealand Clinical Trials Registry (ANZCTR): registration number 12621000348853 (see Appendix E).

4.2.2 Sample size

The sample size computation of 20 was influenced by how many participants could be realistically recruited in the timeframe available and while minimising the disruption to the running of the clinic. It was based on the recruitment numbers in the following studies Bujang et al. (2018), Hackshaw (2008) and Cibere et al. (2008). McNemar's test was used to

compare the test outcomes between the two therapists, with a small sample size this was the maximum precision that was able to be achieved.

4.2.3 Participants

Patients for both studies were recruited from the Counties Manukau District Health Board (CMDHB) outpatient orthopaedic shoulder clinic (OSC). This is a tertiary clinic involved in the assessment of shoulder conditions to triage patients for surgical or conservative management. Patients who attend this clinic have been referred from other specialists or general practitioners, most often due to unsuccessful conservative care or poorly managed pain. If participants had not had an x-ray within six months of their appointment this was completed when they arrived at the clinic as per standard practice.

4.2.4 Inclusion criteria

Potential participants in the studies were required to meet the following inclusion criteria:

- Age ≥18
- Legally able to give consent
- Fluent in English
- Persistent shoulder pain of ≥ 3 months in patients referred to the outpatient shoulder clinic at CMDHB
- Shoulder pain identified as the dominant symptom by both the participant AND the clinician
- Pain of an intensity of ≥2/10 as determined by the Numeric Pain Rating Scale (NPRS) during testing (White et al., 2015)
- A provisional diagnosis of either ACJ or subacromial pain based on the standardised baseline assessment

4.2.5 Exclusion criteria

As part of standard practice for any patient referred to the OSC, potential participants were screened for the following criteria:

- A current or previous history of cancer of the head, neck, thorax or upper limb

- Known rheumatological conditions with musculoskeletal manifestation e.g. polymyalgia rheumatica, spondyloarthropathy or rheumatoid arthritis
- Current or previous osteomyelitis, avascular necrosis, fractures or dislocations around the ipsilateral shoulder complex
- Previous ipsilateral shoulder, thorax or neck surgery
- Previous stroke affecting the ipsilateral upper limb
- Stage five renal failure on dialysis (Davison et al., 2021)
- Any contra-indications to having an anaesthetic injection

This initial screening took place during a phone call made by the researcher or research assistant (BC) (see detail under the 'Recruitment' subheading below).

When the potential participant attended their initial appointment at OSC, further screening was conducted to identify the following exclusion criteria:

- Pain likely to be associated with a frozen shoulder or glenohumeral osteoarthritis (OA) based on their imaging and standardized examination.
    - To rule out frozen shoulder: capsular thickening of more than 2mm on ultrasound or MRI and/or two out of three of the following:
        - >50% loss of passive external rotation compared with the other shoulder (at 90 degrees abduction), hand behind back no higher than the sacrum and less than 100 degrees passive flexion. These criteria were accepted as diagnostic for frozen shoulder (Lewis, 2015; Seo et al., 2012).
    - To rule out glenohumeral OA: a score of ≥3 on the Kellgren Lawrence scale on x-ray
- Ipsilateral upper limb neuropathic pain (e.g. cervical radiculopathy, brachial plexopathy or other peripheral neuropathy) based on their imaging, nerve conduction tests and/or standardized examination.
- Clinical differential diagnoses requiring further investigation (laboratory, other imaging or referral to other medical specialty) i.e. suspicion of metastases, a clinical

history that suggests an undiagnosed inflammatory arthritis or suspicion of an infection.

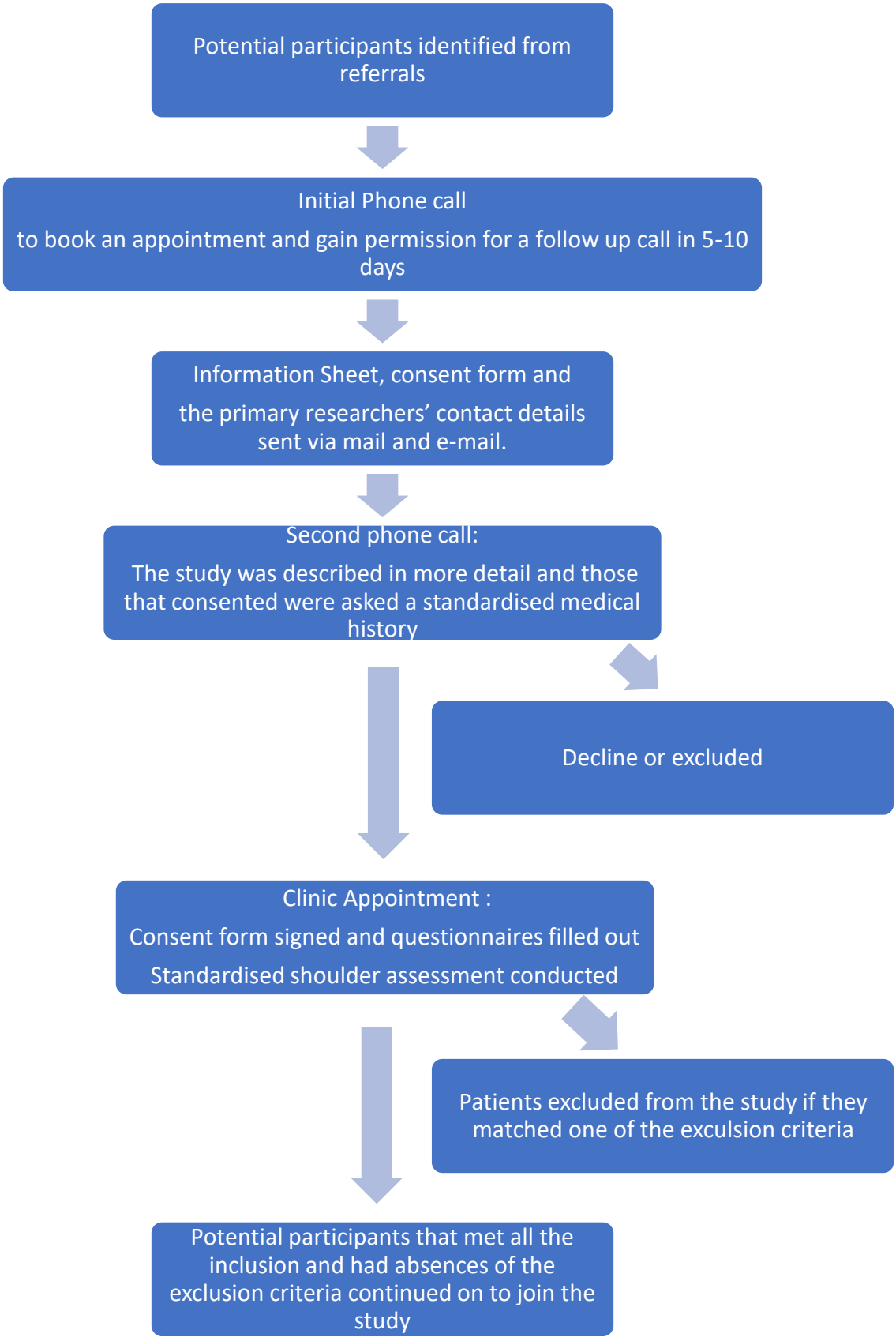
The additional information needed to identify these exclusions was obtained during the standardised baseline assessment.

4.2.6 Recruitment

The first 20 consecutive eligible participants were recruited into the reliability study. Potential participants were identified from those referred to the OSC at the time of triage by the lead researcher and research assistant. Figure 2 provides a visual overview of the recruitment process. Those identified, were contacted by a physiotherapist (the lead researcher or research assistant) to book an appointment time for their clinic consultation. Following this, permission to make a follow-up phone call, between five to 10 days later, was sought. Potential participants were sent the 'Study Information Sheet' (Appendix J), a consent form (Appendix I) and the researchers' contact details via email and mail. This material included a statement to clarify that a decision not to participate in the study would not in any way change the normal assessment and management of their condition provided by the OSC. They were encouraged to phone or email should they wish to ask any questions. The potential participant was also sent a clinic letter that included details of their clinic appointment as per standard practice.

Figure 2

Recruitment Flow Chart



At the time of the follow-up phone call, the potential participant was given a chance to learn more about the research and then to either decline or agree to be involved in the study. Those that agreed, were then asked standardised questions about their current medications, medical history relevant to the shoulder, previous surgery and any previous fractures or dislocations so that those who met any of the relevant exclusion criteria could be identified (see Appendix G). If they took pain medication regularly, they were encouraged to continue this, if they took it intermittently it was requested they not take analgesics in the morning prior to the study's examination or intervention. If they did take their medication however, they were still included if their familiar pain could be reproduced to more than two out of 10 on the NPRS (White et al., 2015). When the potential participant attended their scheduled clinic appointment the research assistant described the study further and gave them another opportunity to have any questions answered before completing a written consent form.

4.2.7 Procedures

Once the consent form was signed participants were asked to fill out a pain diagram, Central Sensitisation Inventory (CSI) and Oxford Shoulder Score (OSS) questionnaires.

Next, they underwent a standardized baseline examination including interview questions and a physical examination (see Appendices O and P). The testing for the reliability study was performed by two physiotherapists, the lead researcher with over 10 years' experience and the other with more than 20 years' experience. The two physiotherapists performing the physical examinations in this study underwent training prior to the commencement of data collection to ensure standardisation of the performance and interpretation of the tests, and familiarity with the study methodology. Clinician order was decided by the toss of a coin and the order of tests was randomised using a computer program to avoid an order effect.

The findings of the first examination enabled the researcher to determine if the potential participant met the study inclusion criteria. Patients who were assessed and found to have a pathology other than ACJ or subacromial pain only went through the first assessment and were then removed from the study. Participants with frozen shoulder, a neck or elbow

pathology could only be identified during the physical examination. Data from this baseline examination for patients who consented to participate but were not eligible was retained for later analysis.

Following the first full baseline examination (interview and physical assessment) the included participant was given a minimum of a 30-minute interval before the second clinician repeated the physical examination component. Prior to each physical examination, the participant was asked to rate their pain intensity at rest. All participants were given standard care management following their second assessment.

Selection of tests

The standardised examination included commonly used questions and tests employed in clinical practice for assessing painful shoulders. The objective was to assess the validity of a range of ACJ and subacromial tests. Given that there are over 70 special orthopaedic tests for the shoulder (Salamh & Lewis, 2020), and that investigating such a large number of tests would require a very large sample size and risk aggravating participants symptoms, a smaller group of well-known and commonly used tests were selected. This selection based on the findings of the literature review reported in Chapter Three and informed by discussions with a number of respected and experienced clinicians working in clinical practice. A patient generated provocative movement (PGPM) was included where the participant was asked to demonstrate their most painful action.

4.2.8 Blinding

The results from the CSI and OSS questionnaires were concealed from the testing clinicians. The first interview and physical assessment results were concealed from the second clinician. This data was not accessible until following the appointment when all the paperwork had been filled out and the potential diagnosis had been reached.

4.2.9 Definition of a positive test

A positive test was defined as a reproduction of the participants' "dominant pain", this was determined by asking them to put one finger over the location of their worst pain. They were also instructed to draw this on the pain diagram with a red cross. Participants during each physical test were asked to point to the area of pain and rate its' severity from zero to

10 points on the NPRS. If their dominant pain was reproduced by two out of 10 or higher, then it was recorded as a positive result (White et al., 2015). Additionally, manual muscle testing (MMT) was included as a measure of strength not pain, this was rated from zero to five on the Oxford scale (Toemen et al., 2011). It was further dichotomised into two groups to improve reliability, these were: severe weakness 0-3/5 or mild weakness 4-5/5.

4.2.10 Statistical analysis

Demographic and other baseline characteristics have been reported as frequencies for categorical data. Continuous data has been presented as means and standard deviations.

Cohen's *κ* two-by-two contingency tables were used to assess inter-rater reliability for each of the 25 tests (De Vries et al., 2008). Cohen's *κ* or kappa is a statistic used to measure the agreement between two quantitative variables (McHugh, 2012). Kappa values however do not adjust for differences in prevalence and bias in the testing population; to overcome this limitation, a Prevalence-Adjusted Bias-Adjusted Kappa (PABAK) was also calculated (Byrt et al., 1993; Chen et al., 2009). Cohen's κ values were interpreted using the following guidelines of Landis and Koch (1977): 0.00-0.20 slight agreement; 0.21-0.40 fair; 0.41-0.60 moderate; 0.61-0.80 substantial; 0.81- 1.00 almost perfect (Landis & Koch, 1977; Newcombe, 1992). The same values have been used to interpret PABAK (Cadogan et al., 2011). The percentage agreement between examiners and the percentage of positive tests was reported (Sullivan et al., 2009). Wilson's method was utilized to calculate their associated confidence intervals (Wilson, 1927).

In addition to comparing the test findings between clinicians, average pain scores across all tests were calculated for each examination. This enabled subsequent statistical analysis to identify any potential significant flare of the participant's pain as a consequence of undergoing multiple tests.

The Mann-Whitney test was utilized to determine if there were significant differences between the demographics and questionnaire scores of those participants included in the study compared to the excluded population (Perme & Manevski, 2019).

A logistic regression was utilized to calculate the importance of the second assessment. The difference between the mean pain scores or odds ratio of pain between the first and second session, adjusting for the identity of the therapist was estimated. P-values were calculated for this to determine whether there was a significant difference (≤0.05) between the testing rounds.

Strength of the rotator cuff was assessed by MMT and a value from zero to five was recorded. The inter-rater agreement, kappa and PABAK were calculated for the strength values. The test outcomes were then grouped to dichotomise the results and assess whether this improved the reliability outcomes.

4.2.11 Data management

The Research Electronic Data Capture (Redcap) software was utilized for data collection, the spreadsheets were designed by the lead researcher (Harris et al., 2009). Dichotomous and discrete data was collected in a flat table format. The participants' data was assigned a unique-number-coded record to conceal their identity. Microsoft Excel was employed to create pivot tables to compare the participants' demographics. Kappa and p-values have been calculated using R version 4.X and *psych* package (R Core Team, 2020; Revelle, 2013).

4.3 Results

The following results have been reported in accordance with the QAREL checklist (Lucas et al., 2013).

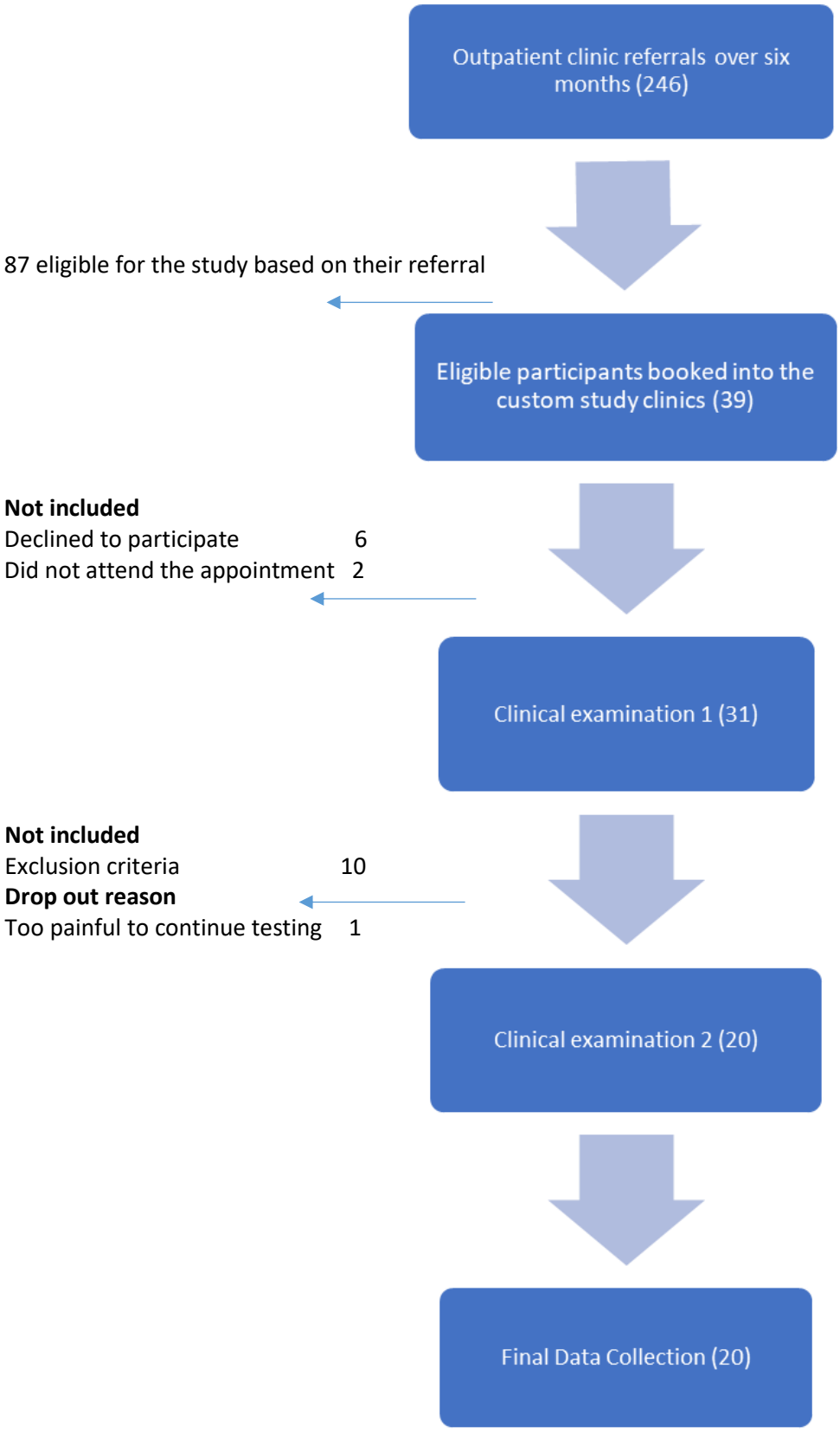
4.3.1 Participants

246 patients referred to the orthopaedic shoulder clinic between March and August 2021 were potential participants. From these referrals 87 patients were eligible based on the information provided. Of the 87 potential participants, 39 were booked into study slots in the clinic, six declined to participate at the second phone call and two did not attend their appointments. Ten participants were excluded based on their clinical assessment and one withdrew due to high levels of pain.

Participants were booked consecutively and once 20 participants had been recruited into the study no further potential participants were included. Figure 3 below outlines the process of recruitment and the reasons for not participating in the study. A current presentation of frozen shoulder was the most common reason for exclusion with four

Figure 3

Flow chart of participants included in the reliability study



patients being excluded for this reason. Two patients had primary neck complaints, one had severe glenohumeral joint osteoarthritis (Kellgren Lawrence Score =5) when x-rayed on arrival, one had a previous neck fusion and another had suffered a previous stroke.

4.3.2 Demographics

Demographic characteristics of the included participants and excluded patients are provided in Table 11 below. Data for the nature of pain and the questionnaire scores were not collected for eight participants in the excluded group, as they either declined or did not attend their appointment. The majority of patients recruited were of European decent, middle aged and male. There were four statistically significant differences between these groups with p-values <0.05. The OSS was 18 in the included participants group compared with 28 in the excluded group (p= 0.004). Additionally, there was a difference in the pain scores of two points on the NPRS from seven in the included participants to nine in the excluded group (p= 0.003). This indicates that the excluded group had a higher perceived shoulder disability and levels of pain (Dawson et al., 2009). There were differences observed in the ethnicities of the participants between the two groups (p= 0.042). Unfortunately, there were higher numbers of Māori and Pasifika patients in the excluded participants as it was our goal to analyse the prevalence data for these groups and to examine ways of improving equity in the clinic through this study. Lastly the included participants were older, with a mean age of 59 compared to 55 in the excluded participants (p= 0.001).

Table 11

Reliability study participant's demographics

	Included	Excluded	P-value
Participants	20	19	-
Number of females	9	6	0.595
Age (years)	59 (12)*	55 (13)*	0.001
Ethnicity			0.042
Asian	3	3	-
European	15	7	-
Pasifika	0	4	-
Māori	2	5	-
Pain duration (months)	21 (26)	59 (119)	0.189
Pain severity (NPRS)	7 (2)*	9 (2)*	0.003
Pain at rest (NPRS)	1 (2)	3 (4)	0.061
CSI	34 (9)	40 (15)	0.421
OSS	28 (7)*	18 (12)*	0.004

Note: Mean values (standard deviation). NPRS: numeric pain rating scale. CSI: Central Sensitisation Inventory. OSS: Oxford Shoulder Score. * Statistically significant differences between the groups.

4.3.3 Reliability of orthopaedic special tests

Table 12 provides detail of the reliability of the OSTs. The mean percentage agreement between the examiners and the percentage of positive results for each test are presented along with the confidence intervals for these figures. The kappa and PABAK values are also displayed for comparison. Missing values in the table reflect where kappa and PABAK could not be calculated due to no overlap of positive results.

Table 12

Percent agreement and reliability of orthopaedic special tests

Tests	Agreement (%)	95% CI		Kappa	95% CI		PABAK	95% CI		Positive Tests (%)	95% CI	
		LB	UB		LB	UB		LB	UB		LB	UB
Wasting triceps	100	86	100	-	-	-	-	-	-	0	0	8
Pseudoparalysis	100	84	100	1.00	1.00	1.00	1.00	0.84	1.00	5	1	16
Wasting supraspinatus	91	73	98	0.62	0.13	1.00	0.83	0.73	0.98	13	6	26
Wasting infraspinatus	91	73	98	0.62	0.13	1.00	0.83	0.73	0.98	13	6	26
Wasting biceps	91	73	98	0.00	0.00	0.00	0.83	0.86	1.00	4	1	15
Wasting deltoid	87	68	95	-0.06	-0.15	0.03	0.74	0.33	0.94	7	2	18
Bilateral ACJ deformity	87	68	95	0.33	-0.25	0.91	0.74	0.33	0.94	11	5	23
Adduction with external rotation	83	61	94	0.56	0.12	1.00	0.67	0.17	0.93	77	62	87
Painful arc	80	58	92	0.00	0.00	0.00	0.60	0.13	0.89	10	4	23
Scapular depression	80	58	92	0.22	-0.33	0.76	0.60	0.13	0.89	15	7	29
Unilateral ACJ deformity	78	58	90	0.16	-0.32	0.62	0.57	0.13	0.85	15	8	28
Scapular elevation	75	53	89	0.29	-0.19	0.77	0.50	0.18	0.83	22	12	37
Bell Van Riet	71	45	88	0.18	-0.36	0.71	0.43	-0.16	0.83	82	67	92
Resisted AC extension	70	48	85	0.34	0.01	0.75	0.40	-0.90	0.76	56	41	70
Scapular protraction	70	49	84	0.25	-0.17	0.68	0.39	-0.60	0.74	28	17	43
Cross arm adduction	68	46	85	0.27	-0.19	0.73	0.37	-0.13	0.75	67	51	79
Paxinos	63	41	81	-0.02	-0.45	0.42	0.26	0.23	0.67	23	12	38
O'Brien's	61	39	80	0.11	-0.33	0.55	0.22	-0.29	0.65	30	18	48
Hawkins Kennedy	58	36	77	0.11	-0.34	0.55	0.16	-0.33	0.60	65	50	78
Scapular retraction	58	36	77	-0.07	-0.48	0.34	0.16	-0.33	0.60	25	14	40
ACJ palpation	55	34	74	0.04	-0.40	0.48	0.10	-0.37	0.54	37	24	52

Note: LB: lower bound. UB: upper bound. CI: confidence interval. PABAK: prevalence-adjusted bias-adjusted kappa. AC: acromioclavicular. ACJ: acromioclavicular joint.

Special tests with a PABAK of less than 0.4 or an inter-rater agreement of less than 70% were considered too poor to be clinically useful (Cadogan et al., 2011; Landis & Koch, 1977; Altman, 1997). The tests that did not attain both of the above criteria were ACJ palpation (PABAK 0.10), scapula retraction (PABAK 0.16), scapula protraction (PABAK 0.39), cross arm adduction (PABAK 0.37), Hawkins Kennedy (PABAK 0.16), O'Brien's (PABAK 0.22) and Paxinos tests (PABAK 0.26). Confidence intervals were not used to determine a threshold of adequate reliability; however lower bounds more than 50% indicate a good probability of agreement.

Pseudoparalysis demonstrated the highest agreement and PABAK values with 100% (CI 95% 84, 100) and 1.00 (CI 95% 0.84, 1.00). Observed wasting of supraspinatus, infraspinatus and the biceps muscle all produced a 91% agreement and a PABAK of 0.83. The PABAK outcomes for these above tests represent a near perfect agreement (Landis & Koch, 1977; Newcombe, 1992). The 95% confidence intervals for supraspinatus and infraspinatus wasting were 0.73-0.98 and 0.86-1.00 for biceps wasting, these both indicate a good probability of agreement. Wasting of the triceps muscle had 100% agreement between raters, the PABAK however could not be calculated as there were zero cases of triceps wasting in the recruited cohort.

Substantial agreement with PABAK scores >0.60 was attained by the following tests, bilateral ACJ deformities with 0.74 (CI 95% 0.33, 0.94), horizontal adduction with external rotation 0.67 (CI 95% 0.17, 0.93) and wasting of the deltoid muscle 0.74 (CI 95% 0.33, 0.94).

4.3.4 Reliability of strength tests

Resisted flexion and abduction did not meet the threshold criteria for percentage agreement or PABAK values (see Table 13). The PABAK outcomes were 0.22 (CI 95% -0.29, 0.65) and 0.20 (CI 95% -0.28, 0.62) respectively. In an effort to improve the reliability of the MMT scores they were grouped into two categories. These were: severe weakness 0-3/5 or mild weakness 4-5/5; see Table 14 for these outcomes. This significantly increased their reliability to 95-100% agreement and a PABAK of 1.00 (CI 95% 0.63, 1.00).

Table 13

Reliability of manual muscle testing

		95% CI			95% CI			95% CI	
Tests	**% Agreement**	LB	UB	**Kappa**	LB	UB	**PABAK**	LB	UB
Flexion	61	39	80	-0.03	-0.48	0.41	0.22	-0.29	0.65
Abduction	60	39	78	0.06	-0.37	0.49	0.20	-0.28	0.62
External rotation	70	48	85	0.37	-1.00	1.00	0.40	-0.09	0.76
Belly press	87	62	96	-0.07	-0.17	0.03	0.73	0.19	0.97

Note: LB: lower bound. UB: upper bound. CI: confidence interval. PABAK: prevalence-adjusted bias-adjusted kappa.

Table 14

Reliability of dichotomised manual muscle testing scores

		95% CI			95% CI			95% CI	
Tests	**% Agreement**	LB	UB	**Kappa**	LB	UB	**PABAK**	LB	UB
Flexion	100	82	100	-	-	-	1.00	0.63	1.00
Abduction	100	83	100	-	-	-	1.00	0.63	1.00
External rotation	95	76	99	0.78	0.67	0.95	0.90	0.50	1.00
Belly press	100	78	100	-	-	-	1.00	0.56	1.00

Note: LB: lower bound. UB: upper bound. CI: confidence interval. PABAK: prevalence-adjusted bias-adjusted kappa.

Dichotomising the test outcomes also resulted in improved reliability for resisted external rotation with a PABAK of 0.90 (CI 95% 0.50, 1.00) and the belly press test with a PABAK of 1.00 (CI 95% 0.56, 1.00).

4.3.5 Percentage of positive outcomes

In addition to collecting data on the reliability of tests, this study also reviewed which tests more commonly reproduced pain ≥2/10. The tests that most commonly provoked pain were the Bell Van Riet and horizontal adduction with external rotation tests. These tests were positive in 82% (95% CI 67, 92) and 77% (95% CI 62, 87) of participants respectively.

The tests least likely to be positive were observed wasting of the deltoid (7%), biceps (4%) or triceps muscles (0%), and pseudoparalysis (5%) which were present in less than 10% of the cohort.

4.3.6 Differences between session one and two

Findings of positive and negative tests between examiner one and two were compared and no statistically significant differences were identified between them. The mean pain scores from the first and second rounds for all OSTs were also analysed. Three tests demonstrated statistically significant differences between the rounds, these were scapular elevation, resisted AC extension and O'Brien's internal rotation ($p \leq 0.05$). Although these tests showed statistically higher pain levels the second time they were tested, they were not clinically significant differences as the pain scores varied by only 0.32/10, 0.14/10 and 0.26/10 respectively. For there to be a clinically significant change in a patient's pain score it must shift at least two points on the scale (White et al., 2015).

4.4 Discussion

This study examined the reliability of shoulder tests that have not previously been reported in the literature in a unique cohort of participants. Poor inter-rater reliability implies that test findings are likely to be variable between clinicians; this may result in different outcomes and ultimately conflicting diagnoses (Apeldoorn et al., 2021). Poor reliability reduces the value of a test, as its' reproducibility determines its' usefulness clinically and ultimately its accuracy (Lange et al., 2017; May et al., 2010). Tests that are not reliable should not be used for diagnostic purposes given that the findings of such tests will be inconsistent.

The majority of tests examined in this study met this study's criteria for adequate reliability of 70% agreement and a 0.40 PABAK. A PABAK of >0.40 is considered to be of moderate reliability (Cadogan, Laslett, et al., 2011a; Landis & Koch, 1977; Newcombe, 1992), this was sufficient to be included in the diagnostic accuracy study. Seven tests in this study demonstrated poor reliability (<0.40) and were not carried forward into the diagnostic accuracy study. These were ACJ palpation, scapula retraction, scapula protraction, cross arm adduction, Hawkins Kennedy, O'Brien's and Paxinos tests.

A number of studies however suggest that a threshold of >0.40 allows for very little agreement and that higher reliability levels are needed for tests to be utilised in the clinical setting (McHugh, 2012). An article by McHugh (2012) and a reliability study by Cadogan, Laslett, et al. (2011a) have both set a minimum acceptable kappa or PABAK threshold at >0.60. Additionally, a level of agreement of 80% or higher has been recommended as a minimum for sufficient interrater agreement in the clinic (Cadogan, Laslett, et al., 2011a; May et al., 2010; McHugh, 2012; Nurjannah & Siwi, 2017).

4.4.1 Strength tests

This study explored the reliability of strength tests given that many authors have reported that, in combination with other tests, weakness is a useful diagnostic tool (Beaudreuil et al., 2009; Cadogan, McNair, Laslett, Hing, et al., 2013; Hegedus et al., 2015; Jain et al., 2017; Park et al., 2005). Few studies to our knowledge have looked at the reliability of MMTs for ACJ and subacromial pathologies. This study therefore investigated the strength outcomes of the MMTs instead of pain provocation. The PABAK outcomes from the MMTs demonstrated that they were only reliable as measures of strength if dichotomised, and this suggested they should not be included as strength tests in the diagnostic accuracy study.

In comparison, a reliability study by Cadogan, Laslett, et al. (2011a) demonstrated strength testing of the rotator cuff can be a reliable measure. They however utilised a hand-held dynamometer instead of MMT, which has been shown to be a more objective and reliable method of measuring strength (Bohannon, 2018; Bohannon, 2019; Cadogan, Laslett, et al., 2011b; Hayes et al., 2002). For pragmatic reasons a dynamometer was not used in this study. The decision was made to not use this device to limit aggravating participant's symptoms with additional tests and to minimise clinic time as they require an average measurement of three attempts.

Furthermore, Lange et al. (2017) in a systematic review and meta-analysis has reported that pain responses to resisted testing can result in moderate to substantial reliability. Additionally, a systematic review and meta-analysis by Lädermann, Collin, et al. (2021) demonstrated that resisted tests have diagnostic value when pain reproduction is the finding of interest. In our study, strength measures with manual muscle testing did not

reach the set reliability threshold. As a result, the decision was made to include MMTs in the diagnostic accuracy study but to record pain responses in lieu of strength outcomes.

4.4.2 Orthopaedic special tests

The seven tests that did not attain both of the above criteria were ACJ palpation, scapula retraction, scapula protraction, cross arm adduction, Hawkins Kennedy, O'Brien's and Paxinos tests. These tests were not carried forward into the diagnostic accuracy study.

The following high performing tests, to our knowledge, have not been investigated in previous reliability studies. These are pseudoparalysis and wasting of the posterior rotator cuff and biceps. The highest performing tests in our study with PABAK outcomes demonstrating near perfect agreement, were an observed wasting of supraspinatus, infraspinatus, and the biceps muscle and pseudoparalysis. This may be explained due to the ease of observing these superficial muscles (Itoi et al., 2006), and that pseudoparalysis was defined as 0°of active elevation with full passive range, which is an obvious pass or fail test (Bauer et al., 2022).

ACJ palpation in this study demonstrated an agreement of 55% (CI 95% 34, 74), a kappa of 0.04 (CI 95% -0.40, 0.48) and a PABAK of 0.10 (CI 95% -0.37, 0.54). This is a test that has not been well researched. To our knowledge there has only been one previous reliability study, by Apeldoorn et al. (2021), and two diagnostic accuracy studies by Walton et al. (2004) and Cadogan, McNair, Laslett and Hing (2013) that have examined this test. Apeldoorn et al. (2021) demonstrated a moderate kappa of 0.47 (CI 95% 0.29, 0.64) for ACJ palpation which is considered poor to moderate depending on the classification criteria used (Cadogan, Laslett, et al., 2011a; Landis & Koch, 1977; Nurjannah & Siwi, 2017). Neither of the diagnostic accuracy studies showed tenderness over the ACJ to be an accurate tool for the diagnosis of ACJ pain, with positive likelihood ratios of 1.07 and 1.37 respectively, and a negative likelihood ratio of 0.87 (Cadogan, McNair, Laslett, & Hing, 2013; Walton et al., 2004). This test however is utilised often in clinical settings and as part of the inclusion criteria in research, despite its' poor to moderate reliability and poor accuracy (Apeldoorn et al., 2021; Chronopoulos et al., 2004; Sirin et al., 2018; van Riet & Bell, 2011; Willimon et al., 2011).

4.4.3 Barriers to comparing reliability studies

The overall results of individual tests in this study demonstrated poor to moderate agreement. This is the same conclusion that Lange et al. (2017) and May et al. (2010) reached on the basis of their systematic reviews on the reliability of physical examination tests for the shoulder. This however is where the similarities end, as there are large variances seen in the quality and reliability outcomes of previous studies that have explored tests for ACJ and subacromial pain (Lange et al., 2017; May et al., 2010). These variations may be explained by differences in the methodologies of reliability research as presented below (Cadogan, Laslett, et al., 2011a; Lange et al., 2017).

Timeframes between the first and second round of testing in reliability studies can be seen to vary significantly. Original studies by Ferenczi et al. (2018) and Razmjou et al. (2004) retested their participants at one week following the first assessment, compared to Burns et al. (2016) who reported five minutes between testing and Nomden et al. (2009) with only a few minutes between. Our study retested participants following 30 minutes to one hour after the first round. A reliability study by Dutil et al. (2017) has recommended the ideal time between testing sessions should be two weeks. This timeframe is thought to minimise any likely changes in symptom presentation or a learning effect. It is however only valid if the condition of interest is stable.

The experience and training of raters in other reliability studies can also be seen to vary considerably; this may influence the outcomes based on the raters' familiarity with the tests and their skill administering them. Vind et al. (2011) reported their examiners had a maximum of six months' clinical experience, compared to Cadogan, Laslett, et al. (2011a) who reported a mean of 29 years' experience. This study in contrast, had two examiners with a mean of 17 years practice. Ericsson et al. (1993) defines expert performance in a profession to be following at least 10 years of learning and growing in that field.

There are significant differences in the demographics of participants included across reliability studies which is beneficial for generalisability, but it can affect the consistency of outcomes (Cadogan, Laslett, et al., 2011a). The previously published reliability studies have recruited from wide ranges of age groups, with acute through to chronic participants from primary through to tertiary care (Apeldoorn et al., 2021; Cadogan, Laslett, et al., 2011a;

Doxey et al., 2018; Johansson & Ivarson, 2009; Nomden et al., 2009; Schmidt et al., 2021; Vind et al., 2011). This study included participants with an average of 59 years of age and a 21 month mean duration of symptoms. The majority of participants reported constant pain and an average of 7/10 pain at worst. The mean CSI results were 34 and the OSS was 28, demonstrating mild sensitisation and moderate to severe symptoms. High pain levels and chronicity have been reported to negatively influence the inter-rater reliability of tests (de Winter et al., 1999; Lange et al., 2017). Our study overall however does not appear to have worse outcomes than the other research in this field.

A further barrier to comparing reliability research is the wide variety of tests assessed. There are over 70 OSTs, and various interpretations of how to perform them, making it difficult to compare between studies (Salamh & Lewis, 2020). The overall consensus in the research is that tests with poor and even moderate reliability should be used sparingly in the clinic, as variable test outcomes can directly influence clinical decision making (Lange et al., 2017; May et al., 2010; Nomden et al., 2009). This is in keeping with the findings of our study.

4.4.4 Comparison of outcomes with the current literature

Only a few studies looking at the reliability of shoulder tests have reported PABAK values which, in contrast to kappa, adjusts for differences in prevalence and bias in the testing population (Cadogan, Laslett, et al., 2011a; Lange et al., 2017). To compare this study's results to the current literature therefore both kappa and PABAK were analysed.

The individual test results produced here differ from the published reliability research. This study has reported a kappa of 0.11 (CI 95% -0.34, 0.55) for the Hawkins Kennedy test, which is significantly lower than that reported by both Apeldoorn et al. (2021) with 0.33 (CI 95% 0.15, 0.51), and Ferenczi et al. (2018) with 0.54 (CI 95% 0.22, 0.80). Lange et al. (2017) in their systematic review and meta-analysis calculated a mean kappa of 0.47 (CI 95% 0.28, 0.67). They concluded that despite the variance seen in reliability their meta-analysis demonstrated a moderate to substantial reliability score for Hawkins Kennedy; our findings however do not support this.

ACJ compression (or O'Brien's test) in our study demonstrated an agreement of 61% (CI 95% 39, 80) and a PABAK of 0.22 (CI 95% -0.29, 0.65). In comparison a reliability study by

Cadogan, Laslett, et al. (2011a) reported an 88% agreement and a 0.75 PABAK. This is a large variation ranging from a weak to a strong agreement (McHugh, 2012). There is one significant difference that may explain the variance in findings between this study and Cadogan, Laslett, et al. (2011a). They reported a mean pain severity of 3.6 out of 10 for their participants, compared to 7.0 in our study; high levels of pain can be shown to confound test results and affect inter-rater agreement (de Winter et al., 1999). A possible explanation for this is participants with high pain levels are likely to be to be aggravated by high volumes of tests. Tests that are later in the sequence of the assessment and/or the second round of testing may be more likely to have positive outcomes for pain provocation, potentially affecting the levels of agreement.

External rotation strength testing demonstrated poor kappa values in our study of 0.37 (95% CI -1.00, 1.00). This was similar to another reliability study by Burns et al. (2016) with a kappa of 0.22 (95% CI 0.00, 0.67). Both our study and Burns et al. (2016) used MMTs to assess muscle strength and ranked the outcomes from 0-5/5. Studies however such as Apeldoorn et al. (2021) and Schmidt et al. (2021) produced higher kappa values of 0.50 (CI 95% 0.34, 0.66) and 0.70 (CI 95% 0.48, 0.92) respectively for resisted external rotation. They also both used MMTs however, Apeldoorn et al. (2021) used pain as an outcome measure not strength and Schmidt et al. (2021) measured strength, but instead of the Oxford scale of 0-5, they recorded whether there was reduced strength compared to the opposite limb (yes or no). These findings emphasise the lack of agreement and uniformity amongst reliability studies and highlight the difficulty interpreting these tests clinically due to such discrepancies in outcomes.

The reliability of individual tests is limited and variable. There may be value however, in grouping results into dichotomous values to improve their validity in the clinic (Lange et al., 2017; May et al., 2010). In our study the reliability of strength testing using MMTs, with the 0-5 grading system of Oxford scoring, was poor. However, we also analysed strength as a discrete variable by dichotomising the results so that all participants were scored yes/no based on whether they had severe weakness (grades 0-3/5) or not. Based on this categorisation, all four resisted tests scored 95-100% agreement and a PABAK of 1.00.

Grouping continuous measures into dichotomous values may have widespread application in improving test reliability.

4.4.5 Implications and clinical application

The ability of any test to diagnose a particular condition depends on both its reliability and diagnostic accuracy. A clinical diagnosis made on the basis of unreliable tests could result in a misdiagnosis and lead to inappropriate treatment (Cadogan, Laslett, et al., 2011a). The majority of OSTs demonstrated a PABAK or kappa <0.60; this is considered to be an inadequate level of agreement by much of the current research (Cadogan, Laslett, et al., 2011a; Lange et al., 2017; May et al., 2010; McHugh, 2012).

As a result of this study's findings the following tests which did not meet the PABAK threshold were not included in the diagnostic accuracy study: ACJ palpation, scapula retraction and protraction, cross-arm adduction, Hawkins Kennedy, O'Brien's and Paxinos tests. In respect to the clinical application of these tests it is recommended that the above tests are interpreted with caution. When utilising tests with variable or poor to moderate reliability the clinician should be aware of the limitations of the tool they are using.

Conversely, our study has demonstrated that wasting of the posterior rotator cuff, biceps and triceps muscles and the presence of pseudoparalysis were highly reliable measures. These tests may be utilised with more confidence, however more research is needed to validate them in the clinical setting, especially in different populations.

4.4.6 Strengths of this study

This study has provided new information to contribute to the known validity of shoulder tests. The reliability of tests such as observed ACJ deformity and muscle wasting, Bell Van Riet, horizontal adduction with external rotation, pseudoparalysis and scapula range of motion have not been assessed previously.

This study demonstrated that in this unique population, with high chronicity and sensitisation, the tests results overall show low to moderate reliability. Our findings in this chronic cohort are similar to previous authors who have looked at less chronic patients (Lange et al., 2017; May et al., 2010). Despite conducting two rounds of 25 tests each on our

participants there did not appear to be any clinically significant differences between the first and second round test outcomes.

4.4.7 Limitations of this study

Whilst this study was designed to meet recommended guidelines for reliability studies, there are some limitations that warrant mentioning.

The confidence intervals around the mean estimates of reliability were reasonably wide due to the relatively small sample size. The mean estimate therefore was not precise, and clinicians should be aware that whilst the mean appears adequate, it might be an over (or under) estimation of the true reliability. The sample size of 20 participants was chosen in order to minimise disruptions in a busy clinic and it was based on the numbers of participants in other similar studies (Bujang et al., 2018; Cibere et al., 2008; Hackshaw, 2008). Maximum precision was able to be achieved using the McNemar's test.

Two examiners with over 10 years of experience performed the testing in this study. It is possible that these tests would be less reliable when employed by less experienced clinicians. This may have affected the generalisability of results and level of agreement.

This study did not include all OSTs for ACJ and subacromial pain; it included 25 commonly used clinical tests in practice for assessing painful shoulders. They were selected based on common tests previously investigated in the literature, the clinical practice of the four experienced clinicians working in the clinic and from using professional networks to canvas other experienced clinicians.

4.5 Conclusion

The aim of this study was to assess the reliability of tests prior to their inclusion in the subsequent diagnostic accuracy study. Fourteen tests showed sufficient reliability to warrant inclusion in this next study. Seven OSTs and strength testing with manual muscle testing were excluded as they did not meet the threshold.

The majority of shoulder tests in this study demonstrated poor to moderate reliability. These findings are consistent with the current literature. The exceptions with high reliability, were pseudoparalysis and wasting of the posterior rotator cuff and biceps.

If a patient's diagnosis, prognosis, and management plan are based on their test outcomes, then it is essential to consider the reliability of the tests utilised. Although most tests showed sufficient reliability to meet the threshold 0.40 PABAK and 70% agreement, this does not mean that they are accurate in identifying or excluding a specific pathology. Adequate reliability is a prerequisite for the validity of a test; the next chapter will examine the diagnostic accuracy of these tests to draw further conclusions on whether they are fit for clinical practice.

Chapter 5 The Diagnostic Accuracy of Shoulder Tests for Subacromial and Acromioclavicular Joint Pain

5.1 Introduction

Many shoulder pathologies have overlapping symptom presentations and very few clinical tests can differentiate between pain sources in this area (Dinnes et al., 2003; Hughes et al., 2008). Physical examination tests as stand-alone tools have been consistently shown to have poor diagnostic accuracy for identifying shoulder conditions (Hegedus et al., 2015; Park et al., 2005; Raynor & Kuhn, 2016). In clinical practice, typically a combination of tests are used along with a patient's history to diagnose a shoulder condition (Cadogan et al., 2012; Raynor & Kuhn, 2016). There is very little published evidence however looking at the utility of combined information from both physical tests and the patient history in the diagnostic process (Cadogan et al., 2012; Raynor & Kuhn, 2016; van Kampen et al., 2014).

The majority of research uses surgery and/or imaging as the reference standard in diagnostic accuracy studies (Cadogan et al., 2012). These methods however do not identify whether the observed pathology is the primary source of pain. Structural changes on MRI, or identified during sugery can be observed in asymptomatic populations (Cadogan, 2011). There is no method to confirm whether they are the primary pain generator (Barreto et al., 2019; Hegedus et al., 2017; Lewis et al., 2015). A marked reduction in pain however following a local anaesthetic injection is considered the most valid reference standard test for the identification of ACJ and subacromial pain (Cardone & Tallia, 2002; McFarland et al., 2017). As well as providing valuable diagnostic data, a positive anaesthetic response can indicate whether a patient may respond to a cortisone injection or even surgical management (Bryceland et al., 2015; Cadogan et al., 2012).

The aim of this study was to determine the diagnostic accuracy of information obtained from the clinical examination of patients with shoulder pain. In particular, the purpose was to determine if any cluster of 'test' findings improved the predictability of a positive anaesthetic response to either an ACJ or subacromial space injection in comparison to information from individual tests.

5.2 Methods

5.2.1 Study design

At the completion of the reliability study, a separate group of potential participants were recruited for the diagnostic accuracy study. They were identified and recruited consecutively following exactly the same procedure as detailed in the previous chapter for the reliability study (see Chapter 4).

In this study, the patient information sheet (see Appendix K) differed from that given to the reliability study participants in that it contained additional details regarding the subacromial +/- ACJ anaesthetic injection (AI).

5.2.2 Sample size

The clinical prevalence data from the recruitment site indicated that approximately 37% of eligible participants were likely to have a diagnosis of ACJ pain. In order to construct an overall predictive model from at most five predictors, considered to be manageable by a clinician (Cowley et al., 2019), a minimum of 50 participants with a positive diagnosis were needed (Harrell, 2015), using the lower end of Harrell's rule and requiring 10 events per predictor. This number of positive diagnoses was predicted to require an overall sample size of 136. With this sample size, specificity and sensitivity for single predictors could be estimated with confidence intervals having expected half-widths of 11 and 15 percentage points respectively (assuming true values of 85% for the specificity and sensitivity). The aim was to recruit 136 potential participants with likely ACJ or subacromial pain.

5.2.3 Data collection

A standardised baseline assessment was performed (see Appendices O and Q) by either the lead researcher (LR) or one of the two orthopaedic surgeons involved in the study. The clinician selection was randomised by flipping a coin. The LR was allocated more clinic slots to book potential participants, therefore a coin was flipped once to select either the LR or a surgeon, if the latter turned up then it was flipped a second time to select the specific surgeon. The order of the physical tests was randomised using a computer randomiser to avoid an order effect.

5.2.4 Reference standard test

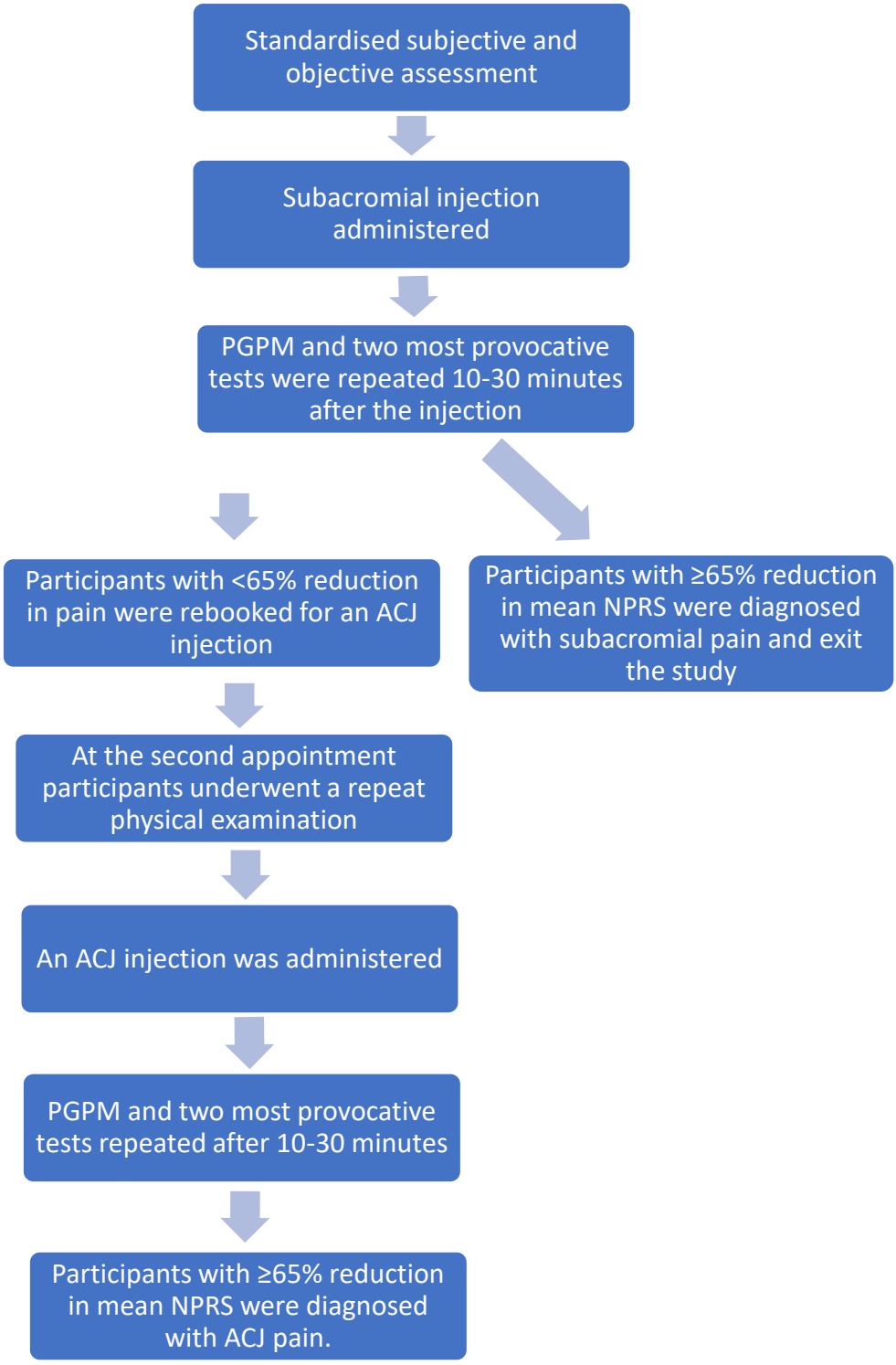
Immediately following the completion of the standardised baseline assessment, the participant was administered a diagnostic subacromial anaesthetic injection (AI). A non-guided AI was administered into the posterior shoulder 2-3 cm below the posterolateral corner of the acromion, in the direction of the coracoid process. The participant was in a seated position with their arm relaxed in their lap. Using a 5ml syringe with a 22-gauge needle, 5ml of lidocaine hydrochloride 1% (xylocaine) was injected by aseptic technique into the subacromial space. The injection was performed by the assessing clinician; either an orthopaedic surgeon or the LR who is a physiotherapist trained to administer such injections under 'standing-orders'.

A mean pain intensity score for each participant was calculated by averaging the reported pain intensity (using the NPRS) of the PGPM and the two most provocative (or first two) of the 18 physical tests (see Appendix Q) performed in the baseline physical examination. The subacromial AI was then administered, and the patient was asked to wait between 10 to 30 minutes before the two most provocative tests from the baseline examination were retested, along with the PGPM. A PAR was considered to be ≥65% reduction in the mean pain intensity score. A PAR was considered evidence that the subacromial space was the source of the participant's pain (Cadogan et al., 2012). This PAR level is consistent with previous thresholds reported in the literature which range between a 50-80% reduction in pain (Cadogan, McNair, Laslett, & Hing, 2013; Datta et al., 2009; Derby et al., 2012; Finniss et al., 2019; Lee et al., 2020; Manchikanti et al., 2020; Strobel et al., 2003; Walton et al., 2004).

Participants who had a PAR to the subacromial injection then exited the study. They were followed up for treatment as per usual practice at the outpatient shoulder clinic. Participants who had <65% reduction in pain intensity following the subacromial injection were considered not to have subacromial pain and hence deemed appropriate for a diagnostic ACJ AI. Those who agreed to undergo this procedure continued in the study and were given a follow up appointment (within 2-4 weeks) at the clinic. See Figure 4 for a detailed outline of the methodology.

Figure 4

Diagnostic accuracy data collection flow chart



Note. ACJ: acromioclavicular joint. NPRS: numeric pain rating scale. PGPM: patient generated painful movement.

At this second appointment, the 18 physical tests from the standardised baseline examination were repeated. This was done by the same clinician who had previously assessed the participant, to establish baseline pain intensity prior to the AI into the ACJ/ACJ capsule. Then an anatomically guided AI was performed by the same clinician (Javed et al., 2017; Strobel et al., 2003). To do this, the joint line was identified by palpating the slight depression between the lateral edge of the clavicle and the acromion. A superior to anterior approach was used from behind, with the participant in a seated position and their arm relaxed in their lap. Using a 5ml syringe with a 22-gauge needle, 3ml of lidocaine hydrochloride 1% (xylocaine) was injected by aseptic technique into and around the ACJ.

Between 10 to 30 minutes after this injection, the PGPM and two most provocative tests were retested by the same clinician and the participant was again required to rate their pain intensity. The mean post AI pain intensity was calculated and compared to the mean pre AI score to record any change in pain. Those participants with ≥65% improvement on the NPRS following this injection were diagnosed with ACJ pain.

5.2.5 Statistical analysis

Demographic and other baseline characteristics were reported as proportions for mean and standard deviation. The included and excluded participants were compared to assess for significant differences between groups. A comparison was also made between the participants with a PAR and those without.

Two by two tables were utilised to calculate various measures of diagnostic accuracy, including specificity, sensitivity, positive and negative likelihood ratios, and positive and negative predictive values. Confidence intervals and p-values were also calculated to establish statistical accuracy. Additive smoothing by adding 0.5 to the 0 value figures was used to enable complete calculations of the likelihood ratios (Kaul et al., 2017).

The accuracy of combinations of test findings was explored via a Firth-penalized logistic regression (Firth, 1993). This is a statistical model that applies a probability function to reduce bias and instability caused by small sample sizes. The coefficients were selected and estimated using the LASSO (least absolute shrinkage and selection operator). A LASSO

regression was utilised as this technique helps to prevent overfitting, and it is capable of dealing with a large number of variables.

The area under the receiver operating characteristic (ROC) curve was maximised using an unweighted model without an intercept. This method was used as it models the relationship between the independent variables (clinical tests) and the dependent variable (a PAR). Models of up to 43 predictors were fitted and evaluated based on the area under the receiver operating characteristic curve (AUC) and misclassification error. To optimise the AUC, 3-fold cross-validation was used after parameter selection and shrinkage through the use of the LASSO (Friedman et al., 2010). To optimise the misclassification error, 'leave one out' cross validation was used. The goal was to obtain a well-performing 5-predictor model, however larger numbers of predictors were also assessed for completeness. Analysis was carried out using the current version of R (Core Team, 2020). The cross-validation and shrinkage was carried out using the *glmnet* package (Friedman et al., 2010).

The regression results refer to a simple regression of a PAR ≥ 65% (true/false). The Firth penalised logistic regression accounts for zero values in the cells, which limit ordinary logistic regressions (Firth, 1993). The estimates and confidence intervals are log-odds-ratios (LOR). A positive LOR corresponds to an increased probability of a PAR ≥ 65% as the predictor increases. A negative LOR corresponds to a decreased probability.

The 43 variables from the interview questions and physical examination were converted into binary values for the two by two tables and receiver operating characteristic curve calculations. A test was considered to be positive if the participants reported pain was >2/10 on the NPRS (White et al., 2015). The OSS outcomes were divided between those with mild to moderate symptoms (≥30), and those with moderate to severe symptoms, ≤29 on the questionnaire scoring. The CSI outcomes were split between those with scores ≥40 which indicates moderate sensitisation or higher and those with <40. Subjective reports of the 'worst pain in the last week' were separated into participants with ≥8/10 pain or those with <8/10; as this was the mean score participants recorded. Similarly, the reported 'lowest pain in the last week' was divided by those with ≥5/10 pain compared to participants with <5/10 pain. This was the mean value of the participants who recorded constant pain.

5.3 Results

The following results have been reported in accordance with the Standards for Reporting of Diagnostic Accuracy studies (STARD) guidelines (Bossuyt et al., 2015).

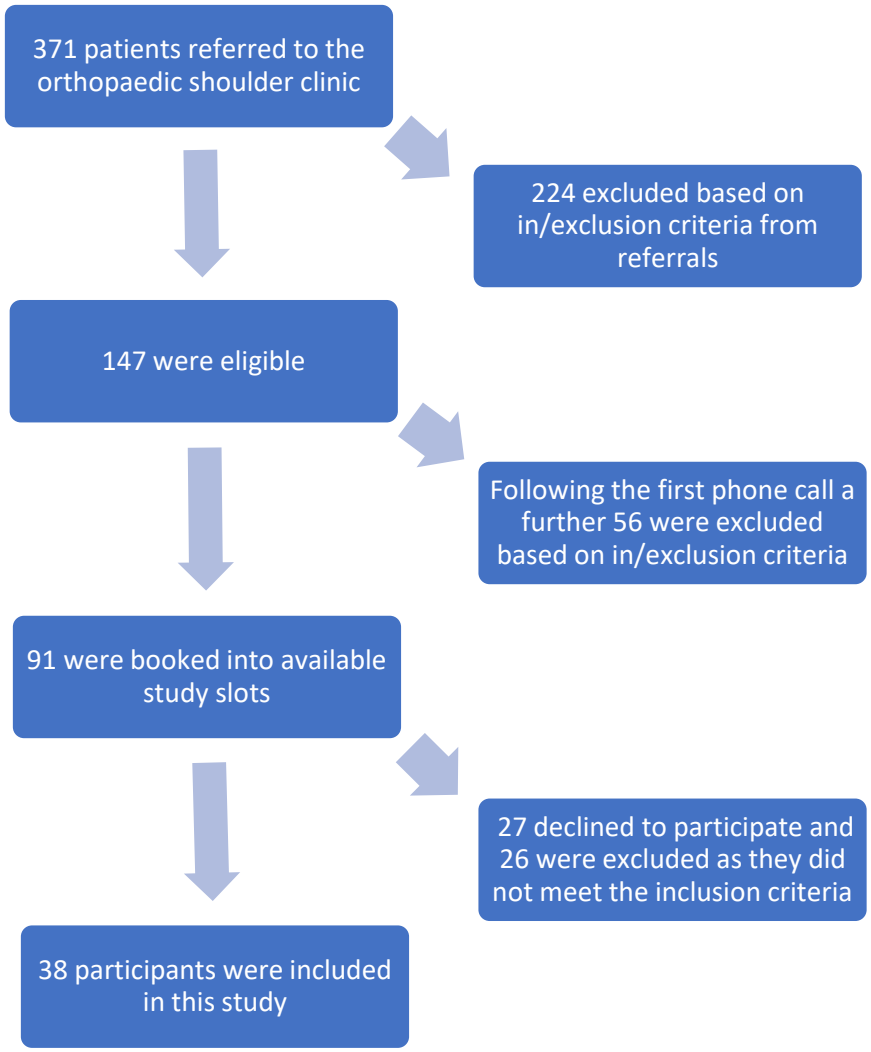
5.3.1 Participants

371 patients referred to the orthopaedic shoulder clinic between April and December 2022 were identified to be potential participants. These referrals were screened and 147 were deemed eligible to participate based on the information provided. Of these, 91 were booked into available study slots to be seen in clinic. The remaining 56 were unable to attend the study times (n=19), no longer required input (n=3), couldn't be reached on the phone (n=5), or were excluded based on inclusion criteria screening over the first phone call (n=29).

Subsequently, of the 91 potential participants booked into the clinic, 27 declined to participate and 26 were excluded as they did not meet the inclusion criteria. This was based on the second phone call or following the initial clinical assessment. As a result, 38 participants were included in this study. This process is outlined in Figure 5 below.

Figure 5

Flow chart of participants included in the diagnostic accuracy study



The main reasons for declining to participate in the study were time constraints and fear of needles. The most common reasons for excluding patients were a diagnosis of frozen shoulder (n=11), difficulty communicating in English (n=6), dominant neck pain (n=2), less than 2/10 pain on the NPRS (n=2) and scapula pain (n=2). Table 15 shows the comparison between the demographics of the included and excluded participants. Data from the participants who declined to participate is not presented here as consent was not obtained.

Table 15

Potential participant's demographics

	Included	Excluded	P-value
Participants	38	26	
Age (years)	56 (13)	59 (13)	0.369
Female	24*	12*	0.005
Ethnicity			0.888
Asian	10	8	-
European	18	10	-
Pasifika	6	4	-
Māori	3	2	-
Middle Eastern/Latin American/African	1	2	-
Pain duration (months)	29 (28)	20 (16)	0.108
Pain severity at worst (NPRS)	8 (1.6)*	6 (3)*	0.004
Pain at rest (NPRS)	2.9 (2.7)*	1.2 (2)*	0.005
CSI	39 (13)	33 (16)	0.119
OSS	24 (9)	26 (10)	0.417

Note: Mean values (standard deviation). NPRS: numeric pain rating scale. CSI: Central Sensitisation Inventory. OSS: Oxford Shoulder Score. * Statistically significant p-value.

There were three statistically significant differences in baseline characteristics between the included and excluded groups when compared with a Mann–Whitney U test calculator. There were 24 females included into the study and 12 females in the excluded group (p= 0.005). Pain levels were higher in the included participants with pain severity at worst (p= 0.004) and pain at rest (p= 0.005) both being two points higher on the NPRS compared to the excluded group.

Both groups reported long durations of symptoms, with the included participants recording an average of 29 months and the excluded 20 months. The mean OSS for the included and excluded participants were 24 and 26 indicating moderate to severe pain. The CSI scores were 39 and 33 respectively for the included and excluded groups, demonstrating mild sensitisation, and the average pain scores at worst were eight and six out ten on the NPRS. These figures indicate that both groups had considerable levels of pain and chronicity.

The overall number of participants was significantly less than the projected number of 136 estimated in the original statistical analysis. Recruitment numbers were low due to feasibility reasons outlined in the discussion. Due to the small numbers the precision estimates were reduced, and there was an increased likelihood of overfitting despite our use of the cross-validation technique.

5.3.2 Demographics

The demographics of participants were analysed to compare those with a PAR to those with a NAR. Only three (8%) of the included participants were Māori and six were Pasifika. These numbers are reflective of the current low levels of clinical engagement in Māori and Pasifika populations in our region (CMDHB, 2019). Efforts were made to address barriers to participation and access to care, including flexible appointment times, reduced wait times, no cost for parking, personal contact by our research team and inclusion of participants whānau (CMDHB, 2019; Graham & Masters-Awatere, 2020). Unfortunately, there were insufficient numbers of Māori and Pasifika participants to allow meaningful comparative statistics to be run.

Thirty-seven of the 38 participants were administered a subacromial injection with one declining. Of these, 14 (38%) experienced a PAR and therefore did not require a follow-up diagnostic injection of the ACJ. Of the 23 participants that had a negative response to the subacromial injection, 10 went on to have the ACJ injection, nine declined a further injection and in four participants there were difficulties accessing the ACJ which made it inappropriate to perform a blind injection, such as adiposity or osteophytes. The one participant that declined a subacromial injection consented to having the ACJ injection; resulting in 11 participants attending the second assessment. Just three (27%) of the ACJ injections resulted in a PAR. Missing data due to human error or incomplete tests due to pain was less than 5%. Demographic data for participants with a PAR, compared to those with a negative anaesthetic response (NAR) can be seen in Table 16 below.

Table 16

Included participant's demographics

	PAR	NAR	P-value
Participants	17	21	-
Number of females	11	14	1.00
Age (years)	57(13)	55 (13)	0.64
Ethnicity	-	-	0.83
Asian	4	6	-
European	9	9	-
Pasifika	2	4	-
Māori	2	1	-
Middle Eastern/Latin American/African	0	1	-
Pain duration (months)	29 (27)	29 (29)	1.00
Pain severity at worst (NPRS)	7.4 (1.8)	8.5 (1.2)	0.04
Pain at rest (NPRS)	3.6 (2.6)	2.3 (2.7)	0.14
CSI	39 (14)	39 (13)	1.00
OSS	25 (8)	23 (9)	0.47

Note: Mean values (standard deviation). PAR: positive anaesthetic response. NAR: negative anaesthetic response. NPRS: numeric pain rating scale. CSI: Central Sensitisation Inventory. OSS: Oxford Shoulder Score.

The Mann Whitney U test was utilised to determine if there were any statistically significant differences between these groups. Only one variable demonstrated a significant difference, pain severity at worst, with a p-value of 0.04, suggesting that participants who reported a NAR had higher pain scores.

5.3.3 Mechanism of injury

Only one variable collected from the participants clinical history demonstrated a statistically significant relationship with a PAR ($p \leq 0.05$). A strain mechanism was statistically associated with a subacromial PAR ($p= 0.04$). The mean estimate of specificity for this mechanism of injury (MOI) was 0.96 (CI 95% 0.78, 1.00) and the positive likelihood ratio (+LR) was 6.57 (CI 95% 0.81, 53.03). The confidence intervals for the +LR crossed one, indicating that this finding was not statistically significant (Hespanhol et al., 2019). Two other tests demonstrated high specificity, these were an insidious onset of pain at 0.78 (CI 95% 0.56, 0.93) and a repetitive MOI at 0.87 (CI 95% 0.66, 0.97). Neither of these had significant +LRs to suggest a clinically useful shift in probability of a PAR when a test is positive.

Tables 17 and 18 below provide detail regarding the diagnostic accuracy of other mechanisms of injury. None of the other mechanisms explored demonstrated a significant relationship with an anaesthetic response. There were three MOIs that did demonstrate a high mean estimate of specificity for an ACJ PAR. These were an 'injury' (external force, fall or impact) and a 'strain mechanism' (stretching, reaching, or lifting), both with a 0.75 (CI 95% 0.35, 0.97) specificity, and an 'insidious onset' at 0.88 (CI 95% 0.47, 1.00). An insidious onset of pain had the highest +LR at 2.67 (CI 95% 0.23, 30.40), however the CIs associated with this figure are wide and cross one.

Table 17

Subacromial mechanism of injury

Variable	TP	FP	FN	TN	Sn (95% CI)	Sp (95% CI)	PPV/ NPV	+LR (95% CI)	-LR (95% CI)	P-value
Injury	5	11	9	12	0.36 (0.13, 0.65)	0.50 (0.31, 0.73)	0.31/0.57	0.75 (0.33, 1.70)	1.23 (0.71, 2.14)	0.47
Insidious	4	5	10	18	0.29 (0.08, 0.58)	0.78 (0.56, 0.93)	0.44/0.64	1.31 (0.42, 4.09)	0.91 (0.61, 1.36)	0.64
Strain	4	1	10	22	0.29 (0.08, 0.58)	0.96 (0.78, 1.00)	0.80/0.69	6.57 (0.81, 53.03)	0.75 (0.53, 1.05)	0.04
Repetitive	4	3	10	20	0.29 (0.08, 0.58)	0.87 (0.66, 0.97)	0.57/0.67	2.19 (0.57, 8.38)	0.82 (0.57, 1.19)	0.24

Note: TP: true positives. FP: false positives. FP: false positives. TN: true negatives. Sn: sensitivity. Sp: specificity. CI: confidence interval. PPV: positive predictive value. NPV: negative predictive value. +LR: positive likelihood ratio. -LR: negative likelihood ratio.

Table 18

Acromioclavicular joint mechanism of injury

Variable	TP[1]	FP	FN	TN	Sn (95% CI)	Sp (95% CI)	PPV/NPV	+LR (95% CI)	-LR (95% CI)	P-value
Injury	1	2	2	6	0.33 (0.00, 0.91)	0.75 (0.35, 0.97)	0.33/0.75	1.33 (0.18, 9.86)	0.89 (0.36, 2.17)	0.78
Insidious	1	1	2	7	0.33 (0.00, 0.91)	0.88 (0.47, 1.00)	0.50/0.78	2.67 (0.23, 30.40)	0.76 (0.33, 1.77)	0.42
Strain	0	2	3	6	0.00 (0.00, 0.71)	0.75 (0.35, 0.97)	0.17/0.65	0.45 (0.03, 7.39)	1.21 (0.70, 2.10)	0.34
Repetitive	1	3	2	5	0.33 (0.00, 0.91)	0.63 (0.24, 0.91)	0.25/0.71	0.89 (0.14, 5.56)	1.07 (0.41, 2.80)	0.90

[1] 0.5 added to zero values when calculating likelihood ratios to calculate outcomes.

Note: TP: true positives. FP: false positives. FP: false positives. TN: true negatives. Sn: sensitivity. Sp: specificity. CI: confidence interval. PPV: positive predictive value. NPV: negative predictive value. +LR: positive likelihood ratio. -LR: negative likelihood ratio.

5.3.4 Pain characteristics

Tables 19 and 20 provide detail regarding the diagnostic accuracy of reported shoulder symptoms. Pain >12 months had a sensitivity of 0.79 (CI 95% 0.49, 0.95) for a subacromial PAR and a -LR of 0.62 (CI 95% 0.20, 1.94). This -LR only indicates a very small, likely insignificant shift in probability. Pain of ≤12 months had a mean estimate specificity of 0.75 (CI 95% 0.35, 0.97) for an ACJ PAR and a +LR of 1.33 (CI 95% 0.18, 9.86). The likelihood ratio and CIs did not reach significance. These statistics indicate that chronicity is not an accurate predictor of a PAR.

The presence of constant pain produced a sensitivity of 0.71 (CI 95% 0.42, 0.92) for a subacromial pain and a -LR of 0.55 (CI 95% 0.22, 1.37). Reported lowest pain levels of ≥5/10 showed a sensitivity of 0.78 (CI 95% 0.56, 0.93) for subacromial pain and a -LR of 0.73 (CI 95% 0.44, 1.21). A -LR of more than 0.5 indicates a very small shift of probability in accurately predicting a subacromial NAR when there is a lowest pain level of ≥5/10. The likelihood ratio CIs also crossed one.

For participants with an ACJ pathology, a lowest pain value of ≥5/10 showed a moderate mean estimate sensitivity of 0.75 (CI 95% 0.35, 0.97). However, the -LR was 1.21 (CI 95% 0.70, 2.10); a -LR of 1 indicates no change in probability of the condition of interest being present or absent on the basis of this test finding.

Table 19

Subacromial pain characteristics

Variable	TP	FP	FN	TN	Sn (95% CI)	Sp (95% CI)	PPV/NPV	LR+ (95% CI)	LR- (95% CI)	P-value
Symptoms for ≤12 months	3	8	11	15	0.21 (0.05, 0.51)	0.65 (0.43, 0.84)	0.27/0.58	0.62 (0.20, 1.94)	1.21 (0.80, 1.81)	0.39
Symptoms for >12 months	11	15	3	8	0.79 (0.49, 0.95)	0.35 (0.16, 0.57)	0.42/0.73	1.21 (0.80, 1.81)	0.62 (0.20, 1.94)	0.39
Worst pain ≥8/10	8	19	6	4	0.57 (0.29, 0.82)	0.17 (0.05, 0.39)	0.30/0.40	0.70 (0.42, 1.13)	2.46 (0.84, 7.23)	0.09
Constant pain	10	11	4	12	0.71 (0.42, 0.92)	0.52 (0.31, 0.73)	0.48/0.75	1.49 (0.87, 2.56)	0.55 (0.22, 1.37)	0.16
Lowest pain ≥5	6	5	8	18	0.43 (0.18, 0.71)	0.78 (0.56, 0.93)	0.55/0.69	1.97 (0.74, 5.27)	0.73 (0.44, 1.21)	0.17
Crepitus	5	9	9	13	0.36 (0.13, 0.65)	0.59 (0.36, 0.79)	0.36/0.59	0.87 (0.37, 2.07)	1.08 (0.64, 1.84)	0.76

Note: TP: true positives. FP: false positives. FP: false positives. TN: true negatives. Sn: sensitivity. Sp: specificity. CI: confidence interval. PPV: positive predictive value. NPV: negative predictive value. +LR: positive likelihood ratio. -LR: negative likelihood ratio.

Table 20

Acromioclavicular joint pain characteristics

Variable	TP[1]	FP	FN	TN	Sn (95% CI)	Sp (95% CI)	PPV/NPV	+LR (95% CI)	-LR (95% CI)	P-value
Symptoms for ≤12 months	1	2	2	6	0.33 (0.00, 0.91)	0.75 (0.35, 0.97)	0.33/0.75	1.33 (0.18, 9.86)	0.89 (0.36, 2.17)	0.78
Symptoms for >12 months	2	6	1	2	0.67 (0.09, 0.99)	0.25 (0.03, 0.65)	0.25/0.67	0.89 (0.36, 2.17)	1.33 (0.18, 9.86)	0.78
Worst pain ≥8/10	1	6	2	2	0.33 (0.00, 0.91)	0.25 (0.03, 0.65)	0.14/0.50	0.44 (0.09, 2.31)	2.67 (0.63, 11.28)	0.20
Constant pain	2	5	1	3	0.67 (0.09, 0.99)	0.38 (0.09, 0.76)	0.29/0.75	1.07 (0.41, 2.80)	0.89 (0.14, 5.56)	0.90
Lowest pain ≥5	0	2	3	6	0.00 (0.00, 0.71)	0.75 (0.35, 0.97)	0.17/0.65	0.45 (0.03, 7.39)	1.21 (0.70, 2.10)	0.34
Crepitus	1	5	1	3	0.50 (0.01, 0.99)	0.38 (0.09, 0.76)	0.17/0.75	0.80 (0.18, 3.54)	1.33 (0.26, 6.94)	0.75

[1] 0.5 added to zero values when calculating likelihood ratios to calculate outcomes.

Note: TP: true positives. FP: false positives. FP: false positives. TN: true negatives. Sn: sensitivity. Sp: specificity. CI: confidence interval. PPV: positive predictive value. NPV: negative predictive value. +LR: positive likelihood ratio. -LR: negative likelihood ratio.

5.3.5 Location of pain

Table 21 and 22 outline the diagnostic accuracy of the reported pain locations. Each of the pain location variables recorded high specificities for both subacromial and ACJ pain with two exceptions. Primary pain above or on the clavicle recorded poor specificity for both conditions and pain radiating below the elbow was not statistically associated with an ACJ PAR.

Pain that radiates below the elbow or into the neck both demonstrated a mean estimate specificity of 0.87 (CI 95% 0.66, 0.97) for subacromial pain. Neither of these variables had +LRs that met statistical significance or demonstrated more than a very small shift in probability. Dominant lateral shoulder pain had a mean estimate of specificity of 0.78 (CI 95% 0.56, 0.93) and a +LR of 2.30 (CI 95% 0.90, 5.86). Pain referred into the hand demonstrated the highest specificity for subacromial pain at 0.96 (CI 95% 0.78, 1.00), however the associated +LR did not reach statistical significance at 0.53 (CI 95% 0.02, 12.26).

Primary pain at the posterior shoulder had a specificity of 1.00 for both a subacromial PAR (CI 95% 0.85, 1.00) and an ACJ PAR (CI 95% 0.63, 1.00). The +LRs were 4.80 (CI 95% 0.21, 110.34) and 2.25 (CI 95% 0.05, 94.61) respectively. Despite the likelihood ratios indicating a small shift in probability, the CIs for both groups crossed one and did not reach statistical significance. The CI's are also very wide inferring that the mean is likely to be far removed from the true value and that more participants were needed to improve the precision of this estimate.

All participants with an ACJ PAR reported their dominant pain to be on or above their clavicle. Whilst this characteristic had a sensitivity of 1.00 (CI 95% 0.29, 1.00), the associated -LR was 0.45 (CI 95% 0.03, 7.39). The CIs were wide and crossed one, indicating no significance in these estimates.

Pain radiating into the neck or hand, lateral or anterior shoulder had a specificity of 0.88 (CI 95% 0.47, 1.00) for an ACJ PAR. The only significant +LR was pain into the hand with 2.67 (CI 95% 0.23, 30.40) demonstrating a small and sometimes important shift in probability. The CIs however crossed one.

Table 21

Subacromial pain location

Variable	TP[1]	FP[1]	FN	TN	Sn (95% CI)	Sp (95% CI)	PPV/ NPV	+LR (95% CI)	-LR (95% CI)	P-value
1° pain: anterior shoulder	3	6	11	17	0.21 (0.05, 0.51)	0.74 (0.52, 0.90)	0.33/ 0.61	0.82 (0.24, 2.77)	1.06 (0.74, 1.53)	0.75
1° pain: posterior shoulder	1	0	13	23	0.07 (0.00, 0.34)	1.00 (0.85, 1.00)	0.75/ 0.64	4.80 (0.21, 110.34)	0.92 (0.77, 1.10)	0.19
1° pain: above/on clavicle	3	11	11	12	0.21 (0.05, 0.51)	0.52 (0.31, 0.73)	0.21/ 0.52	0.45 (30.59, 3.18)	1.51 (0.93, 2.43)	0.11
1° pain: lateral shoulder	7	5	7	18	0.50 (0.23, 0.77)	0.78 (0.56, 0.93)	0.58 /0.72	2.30 (0.90, 5.86)	0.64 (0.36, 1.13)	0.07
2° pain: below elbow	3	3	11	20	0.21 (0.05, 0.15)	0.87 (0.66, 0.97)	0.50/ 0.65	1.64 (0.38, 7.05)	0.90 (0.66, 1.24)	0.50
2° pain: into hand	0	1	14	22	0.00 (0.00, 0.23)	0.96 (0.78, 1.00)	0.25/ 0.61	0.53 (0.02, 12.26)	1.03 (0.90, 1.19)	0.43
2° pain: into neck	2	3	12	20	0.14 (0.02, 0.43)	0.87 (0.66, 0.97)	0.40/ 0.63	1.10 (0.21, 5.77)	0.99 (0.76, 1.29)	0.91

[1] 0.5 added to zero values when calculating likelihood ratios to calculate outcomes.

Note: TP: true positives. FP: false positives. FP: false positives. TN: true negatives. Sn: sensitivity. Sp: specificity. CI: confidence interval. PPV: positive predictive value. NPV: negative predictive value. +LR: positive likelihood ratio. -LR: negative likelihood ratio. 1° pain: primary pain source. 2°pain: secondary pain source.

Table 22

Acromioclavicular pain location

Variable	TP[1]	FP[1]	FN	TN	Sn (95% CI)	Sp (95% CI)	PPV/ NPV	+LR (95% CI)	-LR (95% CI)	P-value
1° pain: anterior shoulder	0	1	3	7	0.00 (0.00, 0.71)	0.88 (0.47, 1.00)	0.25/ 0.68	0.75 (0.04, 14.71)	1.05 (0.66, 1.68)	0.52
1° pain: posterior shoulder	0	0	3	8	0.00 (0.00, 0.71)	1.00 (0.63, 1.00)	0.50/ 0.71	2.25 (0.05, 94.61)	0.93 (0.62, 1.39)	1.00
1° pain: above/on clavicle	3	6	0	2	1.00 (0.29, 1.00)	0.25 (0.03, 0.65)	0.35/ 0.83	1.21 (0.70, 2.10)	0.45 (0.03, 7.39)	0.34
1° pain: lateral shoulder	0	1	3	7	0.00 (0.00, 0.71)	0.88 (0.47, 1.00)	0.25/ 0.68	0.75 (0.04, 14.71)	1.05 (0.66, 1.68)	0.52
2° pain: below elbow	1	3	2	5	0.33 (0.00, 0.91)	0.63 (0.24, 0.91)	0.25/ 0.71	0.89 (0.41, 2.80)	1.07 (0.14, 5.56)	0.90
2° pain: into hand	1	1	2	7	0.33 (0.00, 0.91)	0.88 (0.47, 1.00)	0.50/ 0.78	2.67 (0.23, 30.40)	0.76 (0.33, 1.77)	0.42
2° pain: into neck	0	1	3	7	0.00 (0.00, 0.71)	0.88 (0.47, 1.00)	0.25/ 0.68	0.75 (0.04, 14.71)	1.05 (0.66, 1.68)	0.52

[1] 0.5 added to zero values when calculating likelihood ratios to calculate outcomes.

Note: TP: true positives. FP: false positives. FP: false positives. TN: true negatives. Sn: sensitivity. Sp: specificity. CI: confidence interval. PPV: positive predictive value. NPV: negative predictive value. +LR: positive likelihood ratio. -LR: negative likelihood ratio. 1° pain: primary pain source. 2°pain: secondary pain source.

5.3.6 Functional limitations

The diagnostic accuracy of functional activity limitations for subacromial PAR is reported in Table 23. Three variables demonstrated a sensitivity of 1.00 (CI 95% 0.77, 1.00) for a subacromial PAR; these were: pain reaching behind the back; lying on the affected side and with overhead tasks. The -LRs for these variables were moderate, at 0.18 (CI 95% 0.01, 3.07) for the two former and 0.15 (CI 95% 0.01, 2.45) for the latter. These estimates demonstrate a moderate shift in the probability that should a patient report no trouble reaching their hand up their back, lying on their affected side or with overhead tasks, then a NAR is likely. However, the confidence intervals demonstrate that these estimates are not statistically significant. Not sleeping through the night or having difficulty reaching backwards had a mean estimate sensitivity of 0.86 (CI 95% 0.57, 0.98) for subacromial pain. The -LR for shoulder extension was 0.47 (CI 95% 0.11, 1.95), the CIs however did cross one.

In respect to ACJ pathology, a number of variables demonstrated a sensitivity of 1.00 (CI 95% 0.29, 1.00) as detailed in Table 24. Despite this high sensitivity, none of the associated negative LRs reached statistical significance. The lowest -LR was 0.45 (CI 95% 0.03, 7.39) with painful horizontal adduction, however, the CIs were wide and crossed one.

Table 23

Subacromial pain functional limitations

Variable	TP	FP	FN[1]	TN	Sn (95% CI)	Sp (95% CI)	PPV/ NPV	+LR (95% CI)	-LR (95% CI)	P-value
Pain worse at night-time	6	9	8	14	0.43 (0.18, 0.71)	0.61 (0.39, 0.80)	0.40/ 0.64	1.10 (0.50, 2.42)	0.94 (0.54, 1.64)	0.82
Pain lying on the affected side	14	19	0	4	1.00 (0.77, 1.00)	0.17 (0.05, 0.39)	0.43/ 0.90	1.19 (0.96, 1.47)	0.18 (0.01, 3.07)	0.10
Pain lying on the unaffected side	5	10	9	13	0.36 (0.13, 0.65)	0.56 (0.34, 0.77)	0.33/ 0.59	0.82 (0.35, 1.91)	1.14 (0.67, 1.93)	0.64
Pain disturbed sleep	12	20	2	3	0.86 (0.57, 0.98)	0.13 (0.03, 0.34)	0.38/ 0.60	0.99 (0.76, 1.29)	1.10 (0.21, 5.77)	0.91
Pain with overhead tasks	14	18	0	5	1.00 (0.77, 1.00)	0.22 (0.07, 0.44)	0.44/ 0.92	1.25 (0.99, 1.59)	0.15 (0.01, 2.45)	0.06
Pain reaching up their back	14	19	0	4	1.00 (0.77, 1.00)	0.17 (0.05, 0.39)	0.43/ 0.90	1.19 (0.96, 1.47)	0.18 (0.01, 3.07)	0.10
Painful shoulder extension	12	16	2	7	0.86 (0.57, 0.98)	0.30 (0.13, 0.53)	0.43/ 0.78	1.23 (0.87, 1.74)	0.47 (0.11, 1.95)	0.27
Painful horizontal adduction	10	12	4	11	0.71 (0.42, 0.92)	0.48 (0.27, 0.69)	0.45/ 0.73	1.37 (0.82, 2.29)	0.60 (0.24, 1.52)	0.25

[1] 0.5 added to zero values when calculating likelihood ratios to calculate outcomes.

Note: TP: true positives. FP: false positives. FP: false positives. TN: true negatives. Sn: sensitivity. Sp: specificity. CI: confidence interval. PPV: positive predictive value. NPV: negative predictive value. +LR: positive likelihood ratio. -LR: negative likelihood ratio.

Table 24

Acromioclavicular pain functional limitations

Variable	TP[1]	FP	FN[1]	TN[1]	Sn (95% CI)	Sp (95% CI)	PPV/NPV	+LR (95% CI)	-LR (95% CI)	P-value
Pain worse at night-time	2	3	1	5	0.67 (0.09, 0.99)	0.63 (0.24, 0.91)	0.40/ 0.83	1.78 (0.54, 5.90)	0.53 (0.10, 2.88)	0.39
Pain lying on the affected side	3	7	0	1	1.00 (0.29, 1.00)	0.13 (0.00, 0.52)	0.32/ 0.75	1.05 (0.66, 1.68)	0.75 (0.04, 14.71)	0.52
Pain lying on the unaffected side	0	3	3	5	0.00 (0.00, 0.71)	0.63 (0.24, 0.91)	0.13/ 0.61	0.32 (0.02, 4.88)	1.43 (0.76, 2.71)	0.21
Pain disturbed sleep	3	7	0	1	1.00 (0.29, 1.00)	0.13 (0.00, 0.52)	0.32/ 0.75	1.05 (0.66, 1.68)	0.75 (0.04, 14.71)	0.02
Pain with overhead tasks	3	8	0	0	1.00 (0.29, 1.00)	0.00 (0.00, 0.37)	0.29/ 0.50	0.93 (0.62, 1.39)	2.25 (0.05, 94.61)	1.00
Pain reaching up their back	3	7	0	1	1.00 (0.29, 1.00)	0.13 (0.00, 0.52)	0.32/ 0.75	1.05 (0.66, 1.68)	0.75 (0.04, 14.71)	0.02
Painful shoulder extension	2	5	1	3	0.67 (0.09, 0.99)	0.38 (0.09, 0.76)	0.29/ 0.75	1.07 (0.14, 5.56)	0.89 (0.36, 2.17)	0.90
Painful horizontal adduction	3	6	0	2	1.00 (0.29, 1.00)	0.25 (0.03, 0.65)	0.35/ 0.83	1.21 (0.70, 2.10)	0.45 (0.03, 7.39)	0.34

[1] 0.5 added to zero values when calculating likelihood ratios to calculate outcomes.

Note: TP: true positives. FP: false positives. FP: false positives. TN: true negatives. Sn: sensitivity. Sp: specificity. CI: confidence interval. PPV: positive predictive value. NPV: negative predictive value. +LR: positive likelihood ratio. -LR: negative likelihood ratio.

5.3.7 Questionnaire outcomes

Neither the Central Sensitisation Index or the Oxford Shoulder Score demonstrated any statistically significant correlations with a PAR (see Tables 25 and 26). Although a mean estimate sensitivity of 0.71 (CI 95% 0.42, 0.92) for the OSS was recorded for a subacromial PAR, the -LR of 1.05 (CI 95% 0.36, 3.06) was poor.

The Oxford Shoulder Score demonstrated a mean estimate of sensitivity of 1.00 (CI 95% 0.29, 1.00) for an ACJ PAR, however the associated -LR of 0.45 (CI 95% 0.03, 7.39) was poor and not statistically significant.

Table 25

Subacromial questionnaire outcomes

Variable	TP	FP	FN	TN	Sn (95% CI)	Sp (95% CI)	PPV/ NPV	+LR (95% CI)	-LR (95% CI)	P-value
Oxford score ≤29	10	16	4	6	0.71 (0.42, 0.92)	0.27 (0.11, 0.50)	0.38/ 0.60	0.98 (0.65, 1.49)	1.05 (0.36, 3.06)	0.93
Central sensitisation inventory ≥40	8	13	6	9	0.57 (0.29, 0.82)	0.41 (0.21, 0.64)	0.26/ 0.41	0.97 (0.55, 1.71)	1.05 (0.48, 2.30)	0.91

Note: TP: true positives. FP: false positives. FP: false positives. TN: true negatives. Sn: sensitivity. Sp: specificity. CI: confidence interval. PPV: positive predictive value. NPV: negative predictive value. +LR: positive likelihood ratio. -LR: negative likelihood ratio.

Table 26

Acromioclavicular questionnaire outcomes

Variable	TP	FP	FN[1]	TN	Sn (95% CI)	Sp (95% CI)	PPV/ NPV	+LR (95% CI)	-LR (95% CI)	P-value
Oxford score ≤29	3	6	0	2	1.00 (0.29, 1.00)	0.25 (0.03, 0.65)	0.35/ 0.83	1.21 (0.70, 2.10)	0.45 (0.03, 7.39)	1.00
Central sensitisation inventory ≥40	2	5	1	3	0.67 (0.09, 0.99)	0.38 (0.09, 0.76)	0.29/ 0.75	1.07 (0.41, 2.80)	0.89 (0.14, 5.56)	0.90

[1] 0.5 added to zero values when calculating likelihood ratios to calculate outcomes.

Note: TP: true positives. FP: false positives. FP: false positives. TN: true negatives. Sn: sensitivity. Sp: specificity. CI: confidence interval. PPV: positive predictive value. NPV: negative predictive value. +LR: positive likelihood ratio. -LR: negative likelihood ratio.

5.3.8 Observed deformities

The diagnostic accuracy of observed deformities is presented in Tables 27 and 28 below. Presence of muscle wasting had a mean estimate specificity of 0.96 (CI 95% 0.78, 0.99) and a +LR of 4.93 (CI 95% 0.57, 42.89) for a subacromial PAR. This is a moderate +LR, the CIs however are very wide and cross one. ACJ unilateral or bilateral deformities both had a specificity of 0.83 (CI 95% 0.61, 0.95) and a +LR of 1.64 (CI 95% 0.49, 5.54) for a subacromial diagnosis. This is a very small +LR, indicating a usually unimportant shift in probability in combination with the CIs crossing one.

There were no ACJ PAR participants with muscle wasting, resulting in a mean estimate sensitivity of 1.00 (CI 95% 0.63, 1.00) and a -LR of 0.93 (CI 95% 0.62, 1.39). Bilateral and unilateral ACJ deformity showed a 0.75 (CI 95% 0.35, 0.97) mean estimate of sensitivity and 0.44 -LR (CI 95% 0.09, 2.31) for an ACJ PAR. The CIs once again have crossed one however, which infers the outcomes are statistically insignificant.

Table 27

Subacromial observed deformities

Variable	TP	FP	FN	TN	Sn (95% CI)	Sp (95% CI)	PPV/ NPV	+LR (95% CI)	-LR (95% CI)	P-value
Unilateral ACJ deformity	4	4	10	19	0.29 (0.08, 0.58)	0.83 (0.61, 0.95)	0.50/ 0.86	1.64 (0.49, 5.54)	0.87 (0.59, 1.27)	0.42
Bilateral ACJ deformities	4	4	10	19	0.29 (0.08, 0.58)	0.83 (0.61, 0.95)	0.50/ 0.86	1.64 (0.49, 5.54)	0.87 (0.59, 1.27)	0.42
Muscle wasting	3	1	11	22	0.21 (0.05, 0.51)	0.96 (0.78, 0.99)	0.50/ 0.71	4.93 (0.57, 42.89)	0.82 (0.62, 1.09)	0.10

Note: TP: true positives. FP: false positives. FP: false positives. TN: true negatives. Sn: sensitivity. Sp: specificity. CI: confidence interval. PPV: positive predictive value. NPV: negative predictive value. +LR: positive likelihood ratio. -LR: negative likelihood ratio.

Table 28

Acromioclavicular joint observed deformities

Variable	TP[1]	FP[1]	FN	TN	Sn (95% CI)	Sp (95% CI)	PPV/ NPV	+LR (95% CI)	-LR (95% CI)	P-value
Unilateral ACJ deformity	2	2	1	6	0.67 (0.09, 0.99)	0.75 (0.35, 0.97)	0.50/ 0.86	2.67 (0.63, 11.28)	0.44 (0.09, 2.31)	0.78
Bilateral ACJ deformities	2	2	1	6	0.67 (0.09, 0.99)	0.75 (0.35, 0.97)	0.50/ 0.86	2.67 (0.63, 11.28)	0.44 (0.09, 2.31)	0.78
Muscle wasting	0	0	3	8	0.00 (0.00, 0.71)	1.00 (0.63, 1.00)	0.50/ 0.71	2.25 (0.05, 94.61)	0.93 (0.62, 1.39)	1.00

[1] 0.5 added to zero values when calculating likelihood ratios to calculate outcomes.

Note: TP: true positives. FP: false positives. FP: false positives. TN: true negatives. Sn: sensitivity. Sp: specificity. CI: confidence interval. PPV: positive predictive value. NPV: negative predictive value. +LR: positive likelihood ratio. -LR: negative likelihood ratio.

5.3.9 Active range of motion

Three tests demonstrated a high mean estimate specificity for identifying a subacromial pathology. Presence of a painful arc had a specificity of 0.91 (CI 95% 0.72, 0.99) and +LR of 1.64 (CI 95% 0.26, 10.39) for identifying a subacromial PAR (see Table 29). This LR did not reach statistical significance however as the CIs were poor. Reported pain during active scaption and reaching their hand up their back demonstrated a mean sensitivity of 0.86 (CI 95% 0.57, 0.98) for a subacromial PAR. This however did not reach statistical significance as both -LRs scored higher than 1. Painful active shoulder flexion produced a sensitivity of 0.79 (CI 95% 0.49, 0.95) for a subacromial diagnosis. The -LR was poor with a value of 1.64 (CI 95% 0.38, 7.05).

Active range of motion tests produced no meaningful results for an ACJ PAR with a mean estimate sensitivity of 0.67 (CI 95% 0.09, 0.99) and a specificity of ≤0.25 for all values (see Table 30).

Table 29

Subacromial active range of motion tests

Variable	TP	FP	FN	TN	Sn (95% CI)	Sp (95% CI)	PPV/ NPV	+LR (95% CI)	-LR (95% CI)	P-value
Flexion	11	20	3	3	0.79 (0.49, 0.95)	0.13 (0.03, 0.34)	0.35/ 0.50	0.90 (0.66, 1.24)	1.64 (0.38, 7.05)	0.50
Scaption	12	21	2	2	0.86 (0.57, 0.98)	0.09 (0.01, 0.28)	0.36/ 0.50	0.94 (0.73, 1.20)	1.64 (0.26, 10.39)	0.60
Painful arc	2	2	12	21	0.14 (0.18, 0.43)	0.91 (0.72, 0.99)	0.50/ 0.64	1.64 (0.26, 10.39)	0.94 (0.73, 1.20)	0.60
External rotation	8	19	6	4	0.57 (0.29, 0.82)	0.17 (0.05, 0.39)	0.30/ 0.40	0.69 (0.42, 1.13)	2.46 (0.84, 7.23)	0.09
Reaching up the back	12	21	2	1	0.86 (0.57, 0.98)	0.05 (0.00, 0.23)	0.36/ 0.33	0.90 (0.71, 1.13)	3.14 (0.31, 31.51)	0.30

Note: TP: true positives. FP: false positives. FP: false positives. TN: true negatives. Sn: sensitivity. Sp: specificity. CI: confidence interval. PPV: positive predictive value. NPV: negative predictive value. +LR: positive likelihood ratio. -LR: negative likelihood ratio.

Table 30

Acromioclavicular joint active range of motion tests

Variable	TP	FP	FN	TN[1]	Sn (95% CI)	Sp (95% CI)	PPV/ NPV	+LR (95% CI)	-LR (95% CI)	P-value
Flexion	2	8	1	0	0.67 (0.09, 0.99)	0.00 (0.00, 0.37)	0.23/ 0.25	0.66 (0.31, 1.44)	6.75 (0.34, 132.37)	0.09
Scaption	2	8	1	0	0.67 (0.09, 0.99)	0.00 (0.00, 0.37)	0.23/ 0.25	0.66 (0.31, 1.44)	6.75 (0.34, 132.37)	0.09
External rotation	2	6	1	2	0.67 (0.09, 0.99)	0.25 (0.03, 0.65)	0.25/ 0.67	0.89 (0.36, 2.17)	1.33 (0.18, 9.86)	0.78
Reaching up the back	2	6	1	2	0.67 (0.09, 0.99)	0.25 (0.03, 0.65)	0.25/ 0.67	0.89 (0.36, 2.17)	1.33 (0.18, 9.86)	0.78

[1] 0.5 added to zero values when calculating likelihood ratios to calculate outcomes.

Note: TP: true positives. FP: false positives. FP: false positives. TN: true negatives. Sn: sensitivity. Sp: specificity. CI: confidence interval. PPV: positive predictive value. NPV: negative predictive value. +LR: positive likelihood ratio. -LR: negative likelihood ratio.

5.3.10 Passive range of motion

Table 31 and 32 below, display the diagnostic accuracy of the passive range of motion tests. Both passive flexion and abduction had a mean estimate sensitivity of 0.86 (CI 95% 0.57, 0.98) for a subacromial diagnosis, however only passive abduction demonstrated a -LR less than 0.50. The -LR was 0.47 (CI 95% 0.11, 1.95), however, the CIs crossed one signalling no differences between the PAR and NAR groups.

All passive tests for the ACJ group had a mean estimate sensitivity of 1.00 (CI 95% 0.16, 1.00), and a -LR of 0.60-1.00. These likelihood ratios however were poor, and the CIs crossed one.

Table 31

Subacromial passive range of motion tests

Variable	TP	FP	FN	TN	Sn (95% CI)	Sp (95% CI)	PPV/ NPV	+LR (95% CI)	-LR (95% CI)	P-value
Passive flexion	12	20	2	2	0.86 (0.57, 0.98)	0.09 (0.01, 0.29)	0.38/ 0.50	0.93 (0.73, 1.21)	1.57 (0.25, 9.91)	0.63
Passive abduction	12	16	2	7	0.86 (0.57, 0.98)	0.30 (0.13, 0.53)	0.43/ 0.78	1.23 (0.87, 1.74)	0.47 (0.11, 1.95)	0.27
Passive external rotation	9	19	5	4	0.64 (0.35, 0.87)	0.17 (0.05, 0.39)	0.32/ 0.44	0.78 (0.50, 1.20)	2.05 (0.66, 6.39)	0.21
Passive internal rotation	8	18	6	5	0.57 (0.29, 0.82)	0.23 (0.07, 0.44)	0.31/ 0.45	0.73 (0.44, 1.21)	1.97 (0.74, 5.27)	0.17

Note: TP: true positives. FP: false positives. FP: false positives. TN: true negatives. Sn: sensitivity. Sp: specificity. CI: confidence interval. PPV: positive predictive value. NPV: negative predictive value. +LR: positive likelihood ratio. -LR: negative likelihood ratio.

Table 32

Acromioclavicular joint passive range of motion tests

Variable	TP	FP	FN[1]	TN	Sn (95% CI)	Sp (95% CI)	PPV/ NPV	+LR (95% CI)	-LR (95% CI)	P-value
Passive flexion	2	7	0	1	1.00 (0.16, 1.00)	0.13 (0.00, 0.53)	0.25/ 0.75	1.00 (0.56, 1.79)	1.00 (0.05, 18.57)	0.60
Passive abduction	2	6	0	2	1.00 (0.16, 1.00)	0.25 (0.03, 0.65)	0.28/ 0.83	1.15 (0.60, 2.21)	0.60 (0.04, 9.30)	0.43
Passive external rotation	2	7	0	1	1.00 (0.16, 1.00)	0.13 (0.00, 0.53)	0.25/ 0.75	1.00 (0.56, 1.79)	1.00 (0.05, 18.57)	0.60
Passive internal rotation	2	7	0	1	1.00 (0.16, 1.00)	0.13 (0.00, 0.53)	0.25/ 0.75	1.00 (0.56, 1.79)	1.00 (0.05, 18.57)	0.60

[1] 0.5 added to zero values when calculating likelihood ratios to calculate outcomes.

Note: TP: true positives. FP: false positives. FP: false positives. TN: true negatives. Sn: sensitivity. Sp: specificity. CI: confidence interval. PPV: positive predictive value. NPV: negative predictive value. +LR: positive likelihood ratio. -LR: negative likelihood ratio.

5.3.11 Resisted tests

Pain with resisted testing as individual predictors did not demonstrate a significant relationship with a subacromial anaesthetic response (see Table 33). Resisted flexion demonstrated the lowest values of diagnostic accuracy with a sensitivity of 0.29 (CI 95% 0.08, 0.58) and a specificity of 0.39 (CI 95% 0.20, 0.61). The highest value was the belly press with a specificity of 0.55 (CI 95% 0.27, 0.69).

Conversely all resisted tests scored a mean estimate sensitivity of 1.00 (CI 95% 0.29, 1.00) for predicting an ACJ PAR (see Table 34). The -LRs were between 0.20-0.45, the CIs however did cross one indicating these results are not statistically significant.

Table 33

Subacromial resisted tests

Variable	TP	FP	FN	TN	Sn (95% CI)	Sp (95% CI)	PPV/ NPV	+LR (95% CI)	-LR (95% CI)	P-value
Resisted flexion	4	14	10	9	0.29 (0.08, 0.58)	0.39 (0.20, 0.61)	0.22/ 0.47	0.47 (0.19, 1.14)	1.83 (0.99, 3.35)	0.06
Resisted abduction	5	13	9	10	0.36 (0.13, 0.65)	0.44 (0.23, 0.66)	0.28/ 0.53	0.63 (0.29, 1.39)	1.48 (0.81, 2.72)	0.22
Resisted external rotation	6	13	8	9	0.43 (0.18, 0.71)	0.41 (0.21, 0.64)	0.32/ 0.53	0.73 (0.36, 1.46)	1.40 (0.71, 2.75)	0.34
Belly press	5	12	9	11	0.29 (0.13, 0.65)	0.55 (0.27, 0.69)	0.36/ 0.48	0.65 (0.31, 1.53)	1.28 (0.75, 2.40)	0.33

Note: TP: true positives. FP: false positives. FP: false positives. TN: true negatives. Sn: sensitivity. Sp: specificity. CI: confidence interval. PPV: positive predictive value. NPV: negative predictive value. +LR: positive likelihood ratio. -LR: negative likelihood ratio.

Table 34

Acromioclavicular joint resisted tests

Variable	TP	FP	FN[1]	TN	Sn (95% CI)	Sp (95% CI)	PPV/ NPV	+LR (95% CI)	-LR (95% CI)	P-value
Resisted flexion	3	6	0	2	1.00 (0.29, 1.00)	0.25 (0.03, 0.65)	0.35/ 0.83	1.21 (0.70, 2.10)	0.45 (0.03, 7.39)	0.34
Resisted abduction	3	5	0	3	1.00 (0.29, 1.00)	0.38 (0.09, 0.76)	0.39/ 0.88	1.43 (0.76, 2.71)	0.32 (0.02, 4.88)	0.21
Resisted external rotation	3	3	0	5	1.00 (0.29, 1.00)	0.63 (0.24, 0.91)	0.50/ 0.92	2.25 (0.92, 5.53)	0.20 (0.01, 2.88)	0.21
Belly press	3	4	0	4	1.00 (0.29, 1.00)	0.50 (0.16, 0.84)	0.44/ 0.90	1.75 (0.83, 3.71)	0.25 (0.02, 3.62)	0.12

[1] 0.5 added to zero values when calculating likelihood ratios to calculate outcomes.

Note: TP: true positives. FP: false positives. FP: false positives. TN: true negatives. Sn: sensitivity. Sp: specificity. CI: confidence interval. PPV: positive predictive value. NPV: negative predictive value. +LR: positive likelihood ratio. -LR: negative likelihood ratio.

5.3.12 Orthopaedic special tests

Tables 35 and 36 outline the diagnostic accuracy of the included OSTs. When predicting a subacromial PAR, the highest performing OST was the Bell Van Riet test. It demonstrated a mean estimate sensitivity of 0.92 (CI 95% 0.64, 1.0) and -LR of 0.39 (CI 95% 0.05, 3.07). The CIs however can be seen to cross one.

All three ACJ PAR participants had a positive Bell Van Riet, resisted AC extension and external rotation with horizontal adduction tests. They each scored a mean estimate sensitivity of 1.00 (CI 95% 0.29, 1.00); however resisted AC extension was the only test to produce a low -LR, this was 0.32 (CI 95% 0.02, 4.88). It did not reach statistical significance as the CIs crossed one. Scapula depression demonstrated a specificity of 0.75 (CI 96% 0.35, 0.97) for ruling in an ACJ pathology, and the +LR was 1.33 (CI 95% 0.89, 1.99). The CIs crossed one indicating no statistical difference between the PAR and NAR groups.

Table 35

Subacromial orthopaedic special tests

Variable	TP	FP	FN	TN	Sn (95% CI)	Sp (95% CI)	PPV/ NPV	+LR (95% CI)	-LR (95% CI)	P-value
Scapula elevation	4	8	10	15	0.29 (0.75, 0.02)	0.65 (0.43, 0.84)	0.33/ 0.60	0.82 (0.30, 2.23)	1.10 (0.70, 1.71)	0.70
Scapula depression	3	7	11	16	0.21 (0.05, 0.51)	0.70 (0.47, 0.87)	0.30/ 0.59	0.70 (0.22, 2.29)	1.23 (0.77, 1.66)	0.55
Horizontal adduction in external rotation	7	17	7	5	0.50 (0.23, 0.77)	0.23 (0.08, 0.45)	0.29/ 0.42	0.65 (0.37, 1.15)	2.20 (0.87, 5.59)	0.09
Resisted AC extension	9	13	5	8	0.64 (0.35, 0.87)	0.38 (0.18, 0.62)	0.41/ 0.62	1.04 (0.62, 1.74)	0.94 (0.39, 2.28)	0.89
Bell Van Riet	12	16	1	4	0.92 (0.64, 1.00)	0.20 (0.06, 0.44)	0.43/ 0.80	1.15 (0.88, 1.51)	0.39 (0.05, 3.07)	0.34

Note: TP: true positives. FP: false positives. FP: false positives. TN: true negatives. Sn: sensitivity. Sp: specificity. CI: confidence interval. PPV: positive predictive value. NPV: negative predictive value. +LR: positive likelihood ratio. -LR: negative likelihood ratio. AC: acromioclavicular.

Table 36

Acromioclavicular joint orthopaedic special tests

Variable	TP	FP	FN[1]	TN[1]	Sn (95% CI)	Sp (95% CI)	PPV/ NPV	+LR (95% CI)	-LR (95% CI)	P-value
Scapula elevation	2	4	1	4	0.67 (0.09, 0.99)	0.50 (0.16, 0.84)	0.33/ 0.80	1.33 (0.89, 1.99)	0.67 (0.12, 3.81)	0.62
Scapula depression	1	2	2	6	0.33 (0.00, 0.91)	0.75 (0.35, 0.97)	0.33/ 0.75	1.33 (0.89, 1.99)	0.89 (0.36, 2.17)	0.78
Horizontal adduction in external rotation	3	6	0	1	1.00 (0.29, 1.00)	0.14 (0.00, 0.58)	0.35/ 0.75	1.08 (0.65, 1.77)	0.67 (0.03, 12.96)	1.00
Resisted AC extension	3	5	0	3	1.00 (0.29, 1.00)	0.38 (0.09, 0.76)	0.39/ 0.88	1.43 (0.76, 2.71)	0.32 (0.02, 4.88)	0.21
Bell Van Riet	3	7	0	0	1.00 (0.29, 1.00)	0.00 (0.00, 0.41)	0.32/ 0.50	0.93 (0.62, 1.41)	2.00 (0.05, 83.50)	1.00

[1] 0.5 added to zero values when calculating likelihood ratios to calculate outcomes.

Note: TP: true positives. FP: false positives. FP: false positives. TN: true negatives. Sn: sensitivity. Sp: specificity. CI: confidence interval. PPV: positive predictive value. NPV: negative predictive value. +LR: positive likelihood ratio. -LR: negative likelihood ratio. AC: acromioclavicular.

5.3.13 Multivariate analysis

A multivariate regression analysis, was performed to determine if any combination of variables could predict a subacromial PAR. This analysis could not be performed in respect to the ACJ given the very low numbers of PARs observed within this cohort.

The subacromial PAR coefficients were selected and estimated using a LASSO regression. A total of 43 variables from the interview questions and physical examination were included in the analysis. Variables with continuous outcomes were dichotomised into positive (1) or negative values (0) (see detail in methods page 78). Initially, a cluster of 10 variables was identified as the optimum predictor of a PAR based on the area under curve (AUC). The misclassification error was optimised and produced the same 10 predictor model. The predictors included in this model were: difficulty with overhead tasks, a strain injury onset, lowest pain ≥5/10 (NPRS), presence of muscle wasting, onset of pain from a repetitive activity, worst pain ≥8/10, the primary pain site over or above the clavicle, painful horizontal adduction with external rotation (Add/ER), painful passive internal rotation and painful resisted flexion at 10 degrees. Variables not retained (36 variables) and their coefficients can be found in Appendix U.

When all 10 of the factors were present, the estimated AUC was 0.73 (CI 95% 0.70 ,0.77). This was obtained by using a 3-fold cross-validation technique, to reduce overfitting and to improve the accuracy of the results. The out-of-sample cross-validated misclassification error was calculated to be 35%. This represents a 35% chance of reaching an incorrect outcome when applying the 10 predictors to a new population. The in-sample misclassification error was 13.5%, signifying this model classified participants appropriately 86.5% of the time.

An AUC of one is ideal; comparatively a value of 0.50 (45° diagonal line) indicates a predictive ability no better than guessing 50/50 (Nahm, 2022). For an AUC to be statistically meaningful, it is recommended to be higher than 0.70 to achieve what is considered an acceptable value (Mandrekar, 2010; Terwee et al., 2007).

Table 37 provides further detail of the predictors retained in the model. It demonstrates that some predictors have positive, and others have negative coefficients.

Table 37

LASSO coefficients for subacromial pain: 10 predictor model

Factors	95% CI (LB, UB)	Coefficients	P-values
MOI = strain	0.02, 0.43	0.41	0.05*
MOI = repetitive	-3.60, 0.52	-0.31	0.18
Reported pain at worst ≥8/10	-2.70, 0.23	-0.39	0.10
Reported pain at best ≥5/10	0.06, 3.30	0.37	0.04*
1° pain: above/on clavicle	-2.70, 0.27	-0.56	0.12
Pain with overhead tasks	-0.15, 7.10	0.57	0.07
Painful passive internal rotation	-2.40, 0.45	-0.42	0.18
Painful horizontal adduction in external rotation	-2.90, -0.07	-0.68	0.04*
Painful resisted flexion at 10^0	-2.70, 0.06	-0.45	0.06
Presence of muscle wasting	0.63, 7.90	1.64	0.01*

Note: CI: confidence interval. LB: lower band. UB: upper band. MOI: mechanism of injury. 1°pain: primary pain source. * Statistically significant p-value.

A positive coefficient indicates that if the sign or symptoms are present, there is an increased chance of a positive subacromial PAR. The variables with negative values decrease the probability of a likely subacromial PAR if they are present. The predictors with positive coefficients were: difficulty with overhead tasks, a strain injury onset, lowest pain ≥5/10 (NPRS) and presence of muscle wasting. The negative predictors were: onset of pain from a repetitive activity, worst pain ≥8/10, the primary pain site over or above the clavicle, painful Add/ER, painful passive internal rotation and painful resisted flexion at 10 degrees. Each predictor also had different coefficient values. These values represent the different weighting of importance when considering the prediction of patient outcomes. The presence of muscle wasting, with a coefficient of 1.64 (CI 95% 0.63, 7.90), was the strongest predictor of a subacromial PAR. The predictor with the smallest influence on the probability of a PAR was a repetitive mechanism of injury, this had a coefficient of -0.31 (CI 95% -3.60,

0.52). Confidence intervals and p-values were produced to explore the variables retained. The confidence intervals vary in size, those with a smaller range have an increased certainty in their results. Whilst p-values have been reported, in a predictive setting, they are less useful aside from supporting a true causal relationship between the predictor and the outcome. P-values do not measure the quality of the prediction. The coefficients are more valuable than the p-values in quantifying the contribution of an individual predictor to the overall strength of the model.

To estimate the probability of an individual achieving a PAR, the coefficients corresponding to the positive tests in the predictive model were added to create a score; call it x. The linear score, x, can then be transformed in the equation $p=\frac{e^x}{1+e^x}$ to obtain an estimated probability of a subacromial PAR ≥ 65%. See Table 38 for a participant example and their calculated probability score of having a PAR. In this example, the linear score is -0.24 and the predicted probability score came to 44%. This indicates that there was an estimated 44% probability of a subacromial PAR based on this individual's combined test findings.

Table 38

Participant 1 example of summing LASSO coefficients to obtain linear and probability score

Coefficient factors		Patient results		
Factors	**Coefficients**	**Outcome**		**Score**
MOI = strain	0.41	No	(0)	0.00
MOI = repetitive	-0.31	No	(0)	0.00
Reported pain at worst ≥8/10	-0.39	Yes	(1)	-0.39
Reported pain at best ≥5/10	0.37	No	(0)	0.00
Reported 1° pain: above/on clavicle	-0.56	No	(0)	0.00
Pain with overhead tasks	0.57	Yes	(1)	0.57
Painful passive internal rotation	-0.42	Yes	(1)	-0.42
Painful adduction in external rotation	-0.68	No	(0)	0.00
Painful resisted flexion at 10^0	-0.45	No	(0)	0.00
Presence of muscle wasting	1.64	No	(0)	0.00
Linear score				-0.24
Predicted probability score				44%

Note: MOI: mechanism of injury. 1°pain: primary pain source.

Table 39 provides detail of the probability scores for all participants and the outcome of their anaesthetic response i.e., PAR or NAR. Tables with each of the 37 participant's predicted probability scores can be found in Appendix V.

In Table 39 there are outliers evident that demonstrated a PAR but scored less than 50% probability using this model. Participant one and four both scored a 44% likelihood of having a PAR, however participant one experienced a PAR and number four did not. Each of the participants with a NAR to a subacromial injection scored <44% with this prediction model, with the exception of three participants scoring 44%, 49% and 57%. Twelve of the 14 participants with a PAR scored 44% or higher using this model, despite two misclassified cases with a PAR scoring 24% and 37% predicted probability. This represents a 13.5% in-sample misclassification error. Despite these misclassified cases, should a patient reach a score of ≥60% using this model there is a high likelihood they have a subacromial pathology.

Table 39

Summary of all individual participant's probability scores and anaesthetic response

Participant	Linear score	Predicted probability score	Positive SA PAR
1	-0.24	0.44	yes
2	-0.69	0.33	no
3	-0.24	0.44	yes
4	-0.24	0.44	no
5	-1.24	0.22	no
6	-0.51	0.37	yes
7	-1.24	0.23	no
8	-1.12	0.25	no
9	-0.16	0.46	yes
10	-1.93	0.13	no
11	-0.69	0.33	no
12	-0.09	0.48	yes
13	-0.69	0.33	no
14	-0.69	0.33	no
15	-1.44	0.24	yes
16	-0.13	0.26	no
17	-1.09	0.25	no
18	-1.37	0.20	no
19	0.15	0.54	yes
20	-1.68	0.16	no
21	-0.04	0.49	no
22	0.58	0.64	yes
23	-1.36	0.20	no
24	-0.99	0.27	no
25	1.17	0.76	yes
26	-1.93	0.13	no
27	-0.12	0.53	yes
28	-0.88	0.29	no
29	-1.37	0.20	no
30	-1.93	0.13	no
31	0.63	0.65	yes
32	-1.52	0.18	no
33	0.94	0.72	yes
34	0.27	0.57	no
35	-1.94	0.13	no
36	0.66	0.66	yes
37	2.18	0.90	yes

Note: SA: subacromial. PAR: positive anaesthetic response.

Determining the best AUC by number of predictors

The model presented above contains 10 variables, a number that some might consider too large to practically apply in clinical practice. Given this, the next best model with the lowest predictors and highest AUC associated was sought; this was five with an AUC of 0.65 (CI 95% 0.53, 0.76). Five variables might be considered more manageable in the clinical setting (Cowley et al., 2019). In the five-predictor model, the only positive predictor for a subacromial PAR was the presence of muscle wasting. The negative predictors, which will decrease the overall likelihood of a PAR, were: pain with Add/ER, worst pain ≥8/10, pain bringing their hand up their back and pain with resisted flexion at 10 degrees. See Table 40 for an outline of the variables and their coefficients for this model.

Table 40

LASSO coefficients for subacromial pain: 5 predictor model

Factors	95% CI (LB, UB)	Coefficients	P-values
Reported pain at worst ≥8/10	-2.70, 0.23	-0.14	0.10
Pain reaching their hand up their back	-3.50, 1.10	0.37	0.32
Painful horizontal adduction in external rotation	-2.90, -0.07	-0.33	0.04
Painful resisted flexion at 10^0	-2.70, 0.06	-0.13	0.06
Presence of muscle wasting	0.63, 7.90	0.68	0.01*

Note: CI: confidence interval. LB: lower band. UB: upper band. * Statistically significant p-value.

The 10-predictor model had the highest overall AUC at 0.73, with the lowest number of predictors. The five-predictor model was next the best alternative to this, however with an AUC of 0.65 its predictive power was poor. Much of the research recognises an AUC of 0.60-0.70 to be a poor predictive tool, 0.70 to 0.80 is regarded as fair and a value of ≥0.80 to be excellent (Mandrekar, 2010; Nahm, 2022). A predictive value of 0.73 is considered have an acceptable or moderate discriminatory power to distinguish between variables (Nahm, 2022). This signifies that the 10-predictor model has a moderate power to predict a subacromial PAR, and accordingly a subacromial diagnosis.

5.4 Discussion

5.4.1 Main findings

This study contributes to the current knowledge on the diagnostic accuracy of the clinical examination findings (patient interview and physical tests) as individual and combined variables. It provides further evidence that no individual tool from the clinical examination has sufficient diagnostic accuracy to warrant its' use in ruling in or out an ACJ or subacromial pathology. Additionally, it provides further evidence that demonstrates the importance of considering how combinations of information from the whole clinical examination can guide the clinical diagnosis and decision making. The research was unique in the population recruited, the clinical tests evaluated and the 10-predictor model it has produced.

There is limited research on the diagnostic accuracy of combinations of shoulder tests and even less published on the accuracy of the patient interview (Cadogan et al., 2012; Hegedus et al., 2015; Raynor & Kuhn, 2016). The accuracy of scapula range of motion, the Add/ER test and the combinations of the patient interview questions used in this study have not been investigated previously to our knowledge.

This research has built on the models developed by Cadogan et al. (2012) and Cadogan, McNair, Laslett and Hing (2013) who were the first studies to combine the patient interview and the physical assessment for the diagnostic accuracy of ACJ and subacromial pain, and cast the net wider than OSTs. The addition of our study's findings may aid clinicians to further consider the accuracy of the tests they are using to better inform clinical diagnosis and management.

5.4.2 Participant Cohort

This study has investigated whether commonly utilised clinical examination variables had sufficient diagnostic accuracy to predict an ACJ or subacromial PAR in this unique cohort. Participants were recruited from a large multicultural region that was ethnically and linguistically diverse, with a low socioeconomic status. Additionally, they had high reported chronicity and pain levels. This was an under-represented, minority cohort (Woodland et al., 2021); research in these populations is important to increase knowledge and understanding but also to continue to address health inequities. Recruitment of participants from these

populations however can also be difficult with multiple known barriers (Guerchet et al., 2020).

Investigating the accuracy of the clinical examination in this cohort provides new information. It is difficult to measure the differences between our study's population and cohorts from other similar research; as few articles have documented their participant's duration of symptoms, pain levels or sensitisation (Ackmann et al., 2021; Bak et al., 2010; Collin et al., 2015; Itoi et al., 1999; Walton et al., 2004). Duration of symptoms in those studys that have reported it appear to range from 7 weeks to 21 months (Anauate Nicolao et al., 2022; Cadogan, McNair, Laslett, & Hing, 2013; Heidar Abady et al., 2018; Holtby & Razmjou, 2004; Salaffi et al., 2010). In comparison, the participants in this study reported a mean pain history of 29 months. The average pain score at worst in the study by Cadogan, McNair, Laslett, Hing, et al. (2013) was reported to be 6.2/10, our study demonstrated however a mean value of 8/10 at worst. The average OSS outcome for the included participants in this study was 24, indicating moderate to severe pain and the average CSI score of participants was 39, representing mild sensitisation.

High levels of pain and central-sensitisation can confound the assessment process and negatively impact treatment outcomes (Borstad & Woeste, 2015). The theory is that tests are less accurate in highly painful and sensitised populations. Elevated pain responses can lead to in an increased number of false positive tests due to widespread tenderness in surrounding structures and decreased pain thresholds (Latremoliere & Woolf, 2009).

There is limited research including chronic, sensitised and minority participants in diagnostic accuracy studies for the shoulder. The results in our study demonstrate that that there is no correlation between CSI or Oxford Shoulder Scores and a PAR. Additionally, the overall results for the accuracy of the clinical examination did not differ greatly from the research in different populations (Gismervik et al., 2017; Hegedus et al., 2015; Hughes et al., 2008; Krill et al., 2018). Patients with high levels of pain and/or sensitisation from different backgrounds therefore are no more or less likely to experience a PAR following an AI into either the ACJ or subacromial space.

5.4.3 Utility of the diagnostic injection

A diagnostic AI into the ACJ or subacromial space is considered the reference standard test for these conditions (Cardone & Tallia, 2002; McFarland et al., 2017). In addition to being useful in diagnostic accuracy studies, a PAR may also indicate which patients are likely to respond well to a cortisone injection (Cadogan et al., 2012; Cohen et al., 2013; Schneider et al., 2018). Similarly, a survey of the British Elbow and Shoulder Surgery Society revealed that 77% of respondents viewed a positive response to a subacromial injection as a strong indication of a good surgical outcome (Bryceland et al., 2015). An immediate reduction in pain with shoulder range of motion post injection has also been linked to superior outcomes for patients undergoing a 12-week physiotherapy programme (Marks et al., 2023). The 10-predictor model identified in this research may therefore enhance a clinician's ability to better predict which patient might respond to a subacromial injection and therefore allow for a more timely referral for physiotherapy, a cortisone injection, or surgery.

The threshold for defining a PAR in diagnostic accuracy research varies significantly from a 50-80% reduction in pain (Cadogan, McNair, Laslett, & Hing, 2013; Datta et al., 2009; Derby et al., 2012; Finniss et al., 2019; Lee et al., 2020; Manchikanti et al., 2020; Strobel et al., 2003; Walton et al., 2004). Setting this threshold should take into consideration enhancing the positive predictive value of the injection without excluding participants with the condition. Cohen et al. (2008) performed a study comparing a PAR of 50% to 80%. They concluded that the 80% threshold did not improve outcomes and may result in a missed diagnosis or the exclusion of participants who may benefit from an intervention. In the current study, a handful of participants responded to both the ACJ and subacromial injections, and in one participant the pain reduction was >50% following both injections. A threshold of ≥65% therefore was set to identify the main pain source. This was a pragmatic decision to identify as many true positive cases as possible, taking into account the chronicity and high levels of widespread pain in this cohort (Borstad & Woeste, 2015). By setting the threshold at ≥65%, it is possible the effect of placebo may have influenced the outcomes and a 65% reduction in pain may not have the power to consistently predict treatment outcomes clinically.

5.4.4 The value of orthopaedic special tests as stand-alone diagnostic tests

OSTs are still being taught and utilised in the clinic with the assumption that they can discriminate between different structures in the shoulder and identify a specific pain source (Salamh & Lewis, 2020). Shoulder anatomy however is very complex, with many structures interwoven and continuous with each other (Arai et al., 2014; Clark et al., 1990; Clark & Harryman, 1992; Godenèche et al., 2017; Klatte-Schulz et al., 2022; Roache, 2021; Sahu & Phadnis, 2021). Pain provocation tests for the shoulder are therefore likely to load multiple tissues, making it problematic to identify a specific structure as the source of pain (Salamh & Lewis, 2020). The limitations of OSTs and the presence of asymptomatic abnormalities on imaging can confound the clinical reasoning process (Cadogan et al., 2016).

In this study, OSTs performed poorly as stand-alone clinical tests for the diagnosis of ACJ or subacromial pain in the recruited cohort. The inability for individual shoulder tests to demonstrate a statistically significant correlation with a reference test supports previous research findings (Gismervik et al., 2017; Hegedus et al., 2008; Hegedus et al., 2017; Hughes et al., 2008; Salamh & Lewis, 2020)

Two systematic reviews published in 2008 demonstrated poor outcomes for the accuracy of OSTs (Hegedus et al., 2008; Hughes et al., 2008), since then a number of other systematic reviews have been published with the same outcomes (Alqunaee et al., 2012; Beaudreuil et al., 2009; Biederwolf, 2013; Dakkak et al., 2021; Dinnes et al., 2003; Gismervik et al., 2017; Hanchard et al., 2012; Hegedus et al., 2012; Hermans et al., 2013; Innocenti et al., 2019; Jancuska et al., 2018; Lädermann, Collin, et al., 2021; Lädermann, Meynard, et al., 2021; Papadonikolakis et al., 2011). Hence, there is a large body of research that have disproven the idea that OSTs can identify specific pathologies in the shoulder.

5.4.5 Stand-alone clinical variables for the diagnosis of subacromial pain

No individual variables from the clinical assessment demonstrated a statistically significant relationship with a subacromial anaesthetic response. This conclusion was based on specificity and sensitivity, likelihood ratios, p-values and confidence intervals.

In this study the highest +LR produced was a reported strain (pull, push or lifting) injury at 6.57. The findings of Cadogan et al. (2016) indicated that a history of a strain injury was an accurate tool for identifying a subacromial diagnosis when combined with a range of motion test and reported pain location. Additionally, research by Sørensen et al. (2007) reported that 44% of rotator cuff injuries are brought on by an indirect trauma or strain injury. This may therefore indicate that a history of a strain mechanism of injury could useful when assessing a patient for subacromial pain, in combination with other findings.

The only other moderate +LRs produced were the presence of muscle wasting or dominant pain at the posterior shoulder which had +LRs of 4.93 and 4.80 respectively. Atrophy of the rotator cuff has an established correlation with rotator cuff tears and subacromial pain (Bogdanov et al., 2021; Chung et al., 2013; Laron et al., 2012). It therefore follows that observed muscle wasting of the rotator cuff would be an important variable as part of the wider clinical assessment. To this author's knowledge this is the first study assessing the diagnostic accuracy of observed muscle wasting in the clinical setting.

Itoi et al. (2006) and Gerber et al. (1998) have reported that dominant pain at the posterior shoulder is an uncommon finding for a subacromial pathology. These authors demonstrated that the more common presentations of subacromial pain are the anterior and lateral shoulder regions below the acromion. In the current study, posterior pain was only seen in one participant with a PAR, supporting the findings of Itoi et al. (2006) and Gerber et al. (1998).

The lowest -LR ratios (<0.20) seen in this study for subacromial pain were associated with, pain lying on the affected side, reaching up the back and doing overhead tasks; with values ranging from 0.15 to 0.18. Walton and Russell (2015) have reported that these aggravating factors are the three most commonly described painful movements for symptomatic shoulders with subacromial pain. Whilst our findings were not statistically significant, they support those of these authors and add weight to considering that the absence of these findings might decrease the likelihood of a subacromial diagnosis.

5.4.6 Stand-alone clinical variables for the diagnosis of acromioclavicular pain

Similar to the subacromial outcomes, no single clinical variable produced sufficient levels of accuracy to influence the likelihood of ACJ pain in this cohort. The number of participants with an ACJ PAR was too low to conduct a multivariate analysis.

Pain location and mechanism of injury in this study did not demonstrate sufficient diagnostic accuracy to identify an ACJ pathology. However, although only three participants had a PAR following an ACJ injection, all of them reported their dominant pain as over or above the clavicle and none had a strain injury mechanism. Gerber et al. (1998) injected hypertonic saline into 15 normal ACJs. They reported that the dominant pain pattern pain pattern for the ACJ was straight over the joint and into the upper trapezius-supraspinatus region. Conversely, the pain pattern for a subacromial pathology was described to be below the acromion (Gerber et al., 1998). This is a similar pattern reported by the participants in our study.

In the diagnostic accuracy study by Cadogan, McNair, Laslett and Hing (2013), a repetitive mechanism of pain onset was one of five combined tests shown to significantly increase the likelihood of an ACJ diagnosis. Overhead repetitive activities are known to be a common mechanism of injury for the ACJ (Frantz et al., 2021) and are an important variable to include in future research. A strain mechanism of injury appears to be more associated with a subacromial pathology diagnosis (Cadogan et al., 2016). A study with a larger sample size would be required to further validate whether the presence of a strain injury could significantly decrease the probability of an ACJ diagnosis, as part of the wider clinical assessment.

The findings of this current study support the systematic reviews by Krill et al. (2018) and Hegedus et al. (2008), which have also concluded that no single test is able to demonstrate adequate diagnostic accuracy to rule in or out an ACJ pathology. Although there are less studies for ACJ pain compared to subacromial pain, and only two with high quality methodology (Cadogan, McNair, Laslett, & Hing, 2013; Walton et al., 2004), it does not appear that further research on the accuracy of individual shoulder tests is warranted. Grouping tests together to improve their likelihood ratios has shown to be a more effective method of practice (Cadogan, McNair, Laslett, & Hing, 2013; Chronopoulos et al., 2004; Krill

et al., 2018). The consensus is that combining the patient interview and physical assessment is a more robust method of practice than the reliance on OSTs (Cadogan, McNair, Laslett, & Hing, 2013; Hegedus et al., 2015).

5.4.7 Combining clinical tests

The combination of assessment findings, in a way that more closely reflects clinical practice, resulted in improved diagnostic accuracy for subacromial pain. The highest AUC produced by the LASSO regression with the lowest number of predictors was 10 with an AUC of 0.73. The predictors identified were: difficulty with overhead tasks, a strain injury onset, lowest pain ≥5/10 (NPRS), presence of muscle wasting, onset of pain from a repetitive activity, worst pain ≥8/10, the primary pain site over or above the clavicle, painful Add/ER, painful passive internal rotation and painful resisted flexion at 10 degrees.

An AUC of 0.73 indicates a fair diagnostic accuracy (Nahm, 2022). An AUC of 100 is perfect, an outcome of 0.73 infers that if there were two patients one with a PAR and one with a NAR, then there would be a 73% likelihood that the patient with the PAR would have a higher predictive score. To clarify this further, if the positive variables (signs or symptoms) tested positive in a patient and the negative variables were absent then there would be a high likelihood of a subacromial diagnosis. Because there is no intercept, if none of the predictors are met then the starting probability of a subacromial PAR would be 50%. The out-of-sample misclassification error was 35%, indicating the likelihood of reaching an incorrect outcome when applying the 10 predictors to a new less chronic population. When applying this model to a different clinical population, to be more certain of the outcomes, using a threshold of a predicted probability score of ≥60% is recommended. Furthermore, the optimised misclassification error and AUC calculations both produced the same model; this shows that the predictive model is optimal according to more than one criterion.

Only one of the variables in the 10-predictor model was an OST, this was pain provocation with end range Add/ER. The remaining tests included the mechanism of injury, pain location, pain scores, presence of muscle wasting, range of motion and resisted tests.

The positive variables, indicating an increased likelihood of a subacromial diagnosis if they are present were: difficulty with overhead tasks, a strain injury onset, lowest pain ≥5/10

(NPRS) and presence of muscle wasting. Difficulty with overhead tasks, muscle wasting, and a strain injury mechanism have been identified above as known clinical signs and symptoms associated with subacromial pain and therefore are plausible as robust clinical variables in this model.

Constant pain ≥5/10 was also a positive variable, indicating perhaps that an inflammatory pain pattern is more likely to respond to an AI than a mechanical pain (Yam et al., 2018). This cohort however did have high levels of pain and sensitisation and accordingly there may have been a higher prevalence of constant pain in this population.

The variables with negative coefficient values, which when present imply a decrease in probability of a likely subacromial PAR, were: onset of pain from a repetitive activity, worst pain ≥8/10, the primary pain site over or above the clavicle, painful Add/ER, painful passive internal rotation and painful resisted flexion at 10 degrees.

A handful of these tests traditionally reduce the likelihood of subacromial pain in favour of other pathologies. Positive Add/ER testing, a repetitive mechanism of injury and a reported dominant pain over or above the clavicle may be more likely to indicate an ACJ pathology (Cadogan, McNair, Laslett, & Hing, 2013); therefore, a positive test would help to decrease the likelihood of subacromial pain. Another negative variable was pain with passive internal rotation which, may be associated with a stiff shoulder (Sharma et al., 2015).

The outcomes also suggest that patients with lower pain values may go on to experience a greater reduction in pain post AI. Reported highest pain levels ≥8/10 and painful resisted flexion at 10 degrees were more likely to be associated with a NAR. Clinically however, a patient with subacromial pain would be expected to be symptomatic with resisted testing of the rotator cuff muscles (Varacallo et al., 2023). Participants in this study were instructed to push against the examiners hand only until pain or maximum pressure. This may explain why pain with resisted flexion was a negative predictor, as only pain ≥2/10 was recorded as a positive test.

Patients with subacromial pain often report localised pain around the deltoid region and down to the elbow (Abdelwahab et al., 2021; Gerber et al., 1998). Conversely ACJ pain is

often reported to be above or over the acromion (Gerber et al., 1998). This fits with the findings of pain above or over the clavicle decreasing the likelihood of a subacromial PAR.

Chronopoulos et al. (2004) has demonstrated in their diagnostic accuracy study, that when included in a combination of tests a pain response to compression with horizontal adduction can indicate a symptomatic ACJ. A positive Add/ER test therefore may add value to the wider clinical picture, decreasing the likelihood of a subacromial diagnosis in combination with other variables.

Cadogan et al. (2012), looked at predictors of a PAR from an injection into the subacromial space. They found similar results to our study demonstrating a poor diagnostic accuracy for OSTs and that combined tests are more useful than individual predictors. Two findings obtained from the patient interview (anterior shoulder pain and a strain injury) and one measure of impairment (painfree passive end range external rotation at 90 degrees abduction) produced a specificity of 1.00 for the diagnosis of subacromial pain in their cohort. This outcome supports our study findings where pain below the clavicle, a strain injury and painfree passive range of motion testing is associated with a subacromial pain diagnosis.

The Cadogan, McNair, Laslett, Hing, et al. (2013) diagnostic accuracy study investigated test combinations for medium, large or multi-tendon rotator cuff tears, they also found a cluster of 10 useful tests. Their outcomes were similar to this study with a combination of the patient history and the physical assessment demonstrating statistical significance. The 10 tests included pain with passive range of motion as a negative predictor, strength tests, mechanism of injury and pain characteristics. When five or fewer variables were present the mean estimate sensitivity was 1.00 (CI 95% 0.86, 1.00) and the -LR was 0.00 (CI 95% 0.00, 0.28). When eight or more variables were present the specificity was 0.91 (CI 95% 0.86, 0.95) and +LR 4.66 (CI 95% 2.34, 8.74).

Schmidt et al. (2021) in their study summarised that in the physical assessment range of motion and strength tests appear to be more valid tools than those interpreting pain. Our study only measured pain provocation with resisted and range of motion testing and did not analyse range of motion and strength parameters in the data collection process for

pragmatic reasons. Future research however should include dynamometers and inclinometers to measure these parameters as they have been shown to be a more objective and reliable method of measuring strength and range of motion (Bohannon, 2018; Bohannon, 2019; Cadogan, Laslett, et al., 2011b; Hayes et al., 2002; Tozzo et al., 2021).

Raynor and Kuhn (2016) has published the only systematic review assessing the diagnostic accuracy of the patient history in atraumatic shoulder pain. They found that the combinations of a repetitive strain injury, overhead tasks, age >60, a history hypercholesterolemia and a family history of rotator cuff disease had the highest +LRs for a rotator cuff pathology. None of these clinical variables however reached a +LR >5 (a moderate shift in probability), further research therefore would be needed to validate them.

There is a growing body of evidence, our study included, that indicates combinations of the patient interview, range of motion and strength tests can be accurate tools in combination to diagnose a patient's shoulder complaint (Hegedus et al., 2015; Hegedus et al., 2017; Salamh & Lewis, 2020; Schmidt et al., 2021). These findings represent the early stages in developing a validated model that may be a useful tool in the diagnosis of subacromial pain. The 10 predictor model variables are physiologically plausible and although they are not statistically significant alone, in the model they may be. This model and similar findings by Cadogan et al. (2012) and Cadogan, McNair, Laslett, Hing, et al. (2013) need to be tested in other cohorts and repeated with large sample sizes to validate their utility. It may be difficult to calculate a large number of variables in the clinical setting. Once validated, if a cluster of tests were put into an online calculator that could produce the percentage likelihood of a subacromial diagnosis, however, then this would be a useful tool.

5.4.8 Strengths

This study has built on the existing knowledge of the accuracy of stand-alone clinical tests and investigated their utility in a more chronic and sensitised population. Our study follows the STARD guidelines for reporting and has a clear methodology with well controlled bias. Our outcomes reinforce the understanding of the poor performance of individual clinical shoulder tests and build on the evidence that using combinations of the wider clinical

assessment can enhance the overall diagnostic accuracy. This adds weight to the concept that clinicians are no longer ruling a pathology “in” or “out” but instead combining the assessment outcomes to influence the likelihood of a particular condition being present.

5.4.9 Limitations

The study findings were limited by the size of the sample as indicated by the wide confidence intervals around the mean estimates of accuracy for the majority of tests. The original sample size estimation was for 136 participants to be recruited; however this number was significantly affected by recruiting in a public hospital during the COVID19 pandemic and due to known barriers recruiting from a multicultural population and a low socio-economic region (Guerchet et al., 2020; Woodland et al., 2021).

The utilisation of pain provocation tests to build a differential diagnosis is a less accurate method than utilising more objective outcomes such as measurements of range of motion and/or strength (Hegedus et al., 2015; Hegedus et al., 2017; Raynor & Kuhn, 2016; Salamh & Lewis, 2020; Schmidt et al., 2021). This study recorded reported pain levels with range of motion and resisted testing, and used range of motion as a screening tool for frozen shoulder. The range of motion and strength measurements however were not analysed due to pragmatic reasons. Manual muscle testing and clinician-estimated range of motion were utilised instead of inclinometers or dynamometer; these methods have shown variable validity in the literature (Baschung Pfister et al., 2018). The decision was made to not use these instruments in order to limit participant irritability and minimise clinic time as they require an average measurement of three attempts. There were already 12 range of motion and strength tests in addition to the five special tests included in this study.

There is a consensus in the literature that ultrasound guided shoulder injections are more accurate than using landmark guidance in the clinic (Daniels et al., 2018). The decision to perform the injections in this study without image guidance was a pragmatic one. This is standard practice at the outpatient clinic for subacromial injections, however due to funding, COVID19 and under-staffing it was not possible to align with a radiology team to perform the ACJ guided injections. Research by Kane and Koski (2016) and Strobel et al. (2003) demonstrated that the accuracy rate for a blind (un-guided) subacromial injection is between 70-91%. In comparison, un-guided ACJ injections have shown an accuracy of only

55% (Strobel et al., 2003). Should the un-guided injections in this study have been inaccurate however, participants would be unlikely to have demonstrated a PAR. It is possible though that utilising this technique resulted in some false negative outcomes. Additionally in an ideal set up every participant would have had both injections for comparative data and managing placebo. This was not pragmatic in the busy clinic setting, but also it was a difficult ethical decision to inject a participant who had already responded to a subacromial injection and received their diagnosis.

In this study the reassessment time post injection was set to a minimum of 10 minutes. This was due to the acting time of the anaesthetic (xylocaine) and pragmatic reasons due to restrictions in clinic time and minimising the participant waiting time/impact. The onset of anaesthesia with xylocaine is 5-10 minutes, and it can last from 30 minutes to three hours (Medsafe, 2018; Skinner et al., 1997). Skedros and Pitts (2007) in their study recorded pain relief following subacromial injections at intervals from five to 40 minutes, with a PAR threshold of 75%. They concluded that patients may have enhanced results up to 40 minutes following an injection. Therefore at 10 minutes our cohort may not have reached their peak anaesthetic response and further potential PARs may have been missed.

5.4.10 Clinical implications

This study's findings demonstrate that clinicians should not rely on OSTs or individual clinical variables to inform the diagnostic process. Outcomes from the patient history, resisted tests, observed muscle wasting and range of motion testing should instead be combined to influence the likelihood of a particular condition being present (Salamh & Lewis, 2020). The 10-predictor model produced in this study is a collection of tests that may be of use in the clinical setting to improve the diagnostic process for subacromial pain. This may be more applicable to chronic shoulder pain populations as this was the cohort included in this study.

Our study offers further evidence against the use of orthopaedic special tests for ACJ and subacromial pain diagnostics. An extensive amount of literature has been published over the last two decades that demonstrates significant limitations for their utility in the clinical setting (Hegedus et al., 2017).

5.4.11 Future research

This research supports the current evidence on shoulder diagnostics and more specifically orthopaedic special tests. It has confirmed that in a chronic and sensitised cohort these tests do not add diagnostic value individually in the clinic. Further research in this space analysing individual clinical tests for their diagnostic accuracy is not justified. It is clear a diagnosis should be based on the whole picture clinically, not on the findings of individual tests, and research going forward needs to reflect this.

The 10-predictor model introduced here was preliminary and produced from low recruitment numbers; therefore further prospective validation with larger numbers in different populations is required. Future research should follow a standardised framework for reporting and use anaesthetic injections as a reference standard.

There is a need for a clinical guideline to be produced outlining the most valid use and interpretation of shoulder assessment tools to standardise the diagnostic process. Ultimately an online calculator would simplify the application of future validated diagnostic models to facilitate the diagnosis of shoulder complaints across the healthcare profession.

5.5 Conclusion

The traditional process of diagnosing a shoulder pathology has been to rule a pathology in or out a based on the findings of “special tests”. These tests however, have been disproven by a large body of research and the current use of them in practice needs to change.

Our study’s findings support the current research that OSTs or any stand-alone tests are not able to identify an ACJ or subacromial diagnoses. In contrast, this research suggests that a combination of findings from the wider clinical assessment can help to sufficiently influence the likelihood of diagnosing these pathologies. A 10 predictor model has been created here which may be useful clinically in the diagnosis of chronic subacromial pain.

These findings are specific to this cohort with chronic shoulder pain. Further research is required to determine if the results are reproducible and if they might be generalisable to patients with less severe symptoms. The focus should be on combinations of the wider clinical assessment along with diagnostic injections as a reference test.

Chapter 6 Summary, Clinical Implications and Conclusions

This chapter will summarize the key findings from the reliability and diagnostic accuracy studies and the systematic review. The conclusions from these original studies will be compared to the current research. The outcomes likely to influence clinical practice will be discussed in addition to the recommendations for future research. The main findings from this thesis are below.

1. The majority of shoulder tests have poor to moderate inter-rater reliability.

2. Individual shoulder tests or clinical variables have poor diagnostic accuracy and do not create any meaningful shift in likelihood of an ACJ or subacromial diagnosis.

3. A combination of variables from the wider clinical assessment including the patient interview, clinician observations and resisted and range of motion tests are more useful tools than orthopaedic special tests. A multivariate analysis identified a specific combination of ten clinical variables that improved the prediction of subacromial pain in this cohort.

6.1 Key findings

6.1.1 Reliability of clinical tests for acromioclavicular and subacromial pain

The reliability study assessed 25 commonly utilised tests for ACJ and subacromial pathologies to determine their reliability prior to including them in the diagnostic accuracy study. Fourteen tests met the threshold of ≥0.40 PABAK and ≥70% agreement and were included in the subsequent study. The reliability of seven tests were too poor to be included (<0.4 PABAK); these were ACJ palpation, scapula retraction and protraction, cross-arm adduction, Hawkins Kennedy, O'Brien's and Paxinos tests.

The majority of these tests demonstrated poor to moderate inter-rater reliability. The exception to this, with high reliability, was pseudoparalysis and observed posterior rotator cuff and biceps muscle wasting. In addition to the OSTs, the reliability of strength outcomes with manual muscle testing was also assessed. The tests using the Oxford scale demonstrated poor reliability, however when grouped into dichotomous categories of severe weakness (0-3/5) and mild weakness (4-5/5) they had a near perfect agreement.

Although most tests showed sufficient reliability to meet the threshold set, this does not imply that this level of agreement is adequate for clinical practice. A number of studies have recommended a level of agreement of ≥80% and a kappa or PABAK of ≥0.60 to be the minimum for sufficient interrater agreement in the clinic (Cadogan, Laslett, et al., 2011a; May et al., 2010; McHugh, 2012; Nurjannah & Siwi, 2017). It has been suggested that any reliability outcomes lower than this allows for very little agreement and tests that do not meet these levels should not be utilised in the clinical setting (McHugh, 2012).

Our reliability study findings of poor to moderate inter-rater agreement were consistent with the outcomes reached in our systematic review (see Chapter 3). The systematic review concluded that across all the of OSTs only the belly press and internal rotation lag sign showed high inter-rater reliability. There were also wide variations in the methodology and outcomes of reliability studies which made it difficult to draw comparisons and reach conclusions (Lange et al., 2017). Our review and the systematic reviews by Lange et al. (2017) and May et al. (2010) have presented both high and low kappa/PABAK scores for the same test across multiple studies. This calls into question whether clinical tests, especially OSTs can be relied on in practice.

Clinical tests with poor reliability produce inconsistent results and consequently, this may result in incorrect interpretations of the clinical assessment and a misdiagnosis for the patient (Cadogan, Laslett, et al., 2011a). Tests require both adequate reliability and diagnostic accuracy to be useful tools in a clinical setting (Apeldoorn et al., 2021).

6.1.2 Diagnostic accuracy of clinical tests for acromioclavicular and subacromial pain

Our diagnostic accuracy study was unique due to the clinical information examined and the chronic, sensitised cohort recruited. The results from our diagnostic accuracy study and systematic review demonstrated that all clinical variables, including OSTs, have limited accuracy when utilised as stand-alone tests for the diagnosis of ACJ or subacromial pain (Hegedus et al., 2017; Salamh & Lewis, 2020). Combining the clinical variables from the wider assessment, however, can result in adequate diagnostic accuracy to influence the likelihood of an ACJ or subacromial diagnosis (Cadogan et al., 2012; Cadogan, McNair, Laslett, & Hing, 2013).

The diagnostic accuracy study presented a predictive model of 10 variables with an AUC of 0.73 for diagnosis of subacromial pain. There were however too few numbers of PARs to conduct a similar multivariate analysis for the ACJ.

The variables in the 10-predictor model with positive coefficients were: difficulty with overhead tasks, a strain injury onset, lowest pain ≥5/10 (NPRS) and presence of muscle wasting. This infers that if these variables are positive in a patient, there is an increased likelihood of a subacromial diagnosis. The other variables had negative coefficients, therefore if the signs or symptoms were present then they would decrease the likelihood of a subacromial diagnosis. The inverse is also true, for example if the negative variables tested negative then they would increase the possibility of a subacromial diagnosis. The negative predictors were: onset of pain from a repetitive activity, worst pain ≥8/10, the primary pain site over or above the clavicle, painful horizontal adduction with external rotation, painful passive internal rotation and painful resisted flexion at 10 degrees.

The clinical variables in the model had different coefficient values representing the different weighting of importance when predicting the presence of subacromial pain. The presence of muscle wasting, was the strongest predictor and a repetitive mechanism of injury had the smallest influence on the probability of a subacromial PAR.

This model classified participants appropriately 86.5% of the time within the sample. Should this cluster of clinical tests be applied to a new population there is a chance of a 35% misclassification error. Clinically 10 variables would be difficult to calculate; however if the test results were entered into an online calculator that produced the predicted probability of a subacromial diagnosis this would be a practical tool.

6.2 Contributions from this thesis

No systematic reviews have addressed both the reliability and diagnostic accuracy of shoulder tests for ACJ and subacromial pain. This research has contributed useful data to the current knowledge of the clinical tools used to diagnose these shoulder pathologies. There is little known about the inter-rater reliability of some of the tests included in the reliability study. To date, few researchers have examined the diagnostic accuracy of specific combinations of shoulder tests and to our knowledge none have considered the utility of

combinations of findings obtained from the patient interview and the physical examination for identifying both ACJ and subacromial pathologies. Our findings add weight to the knowledge that specific combinations of information obtained from both the history and physical examination are better at predicting an ACJ or subacromial diagnosis than individual tests (Cadogan et al., 2012; Cadogan, McNair, Laslett, & Hing, 2013; Salamh & Lewis, 2020).

6.3 The importance of a pathoanatomical diagnosis

There is a debate in the literature over whether a specific pathoanatomical diagnosis needs to be made or whether 'non-specific shoulder pain' is an appropriate umbrella term to use in clinical practice (Hegedus et al., 2017; Klintberg et al., 2015). An accurate pathoanatomical diagnosis may inform a patient about their prognosis, but also guide decision making in respect to conservative versus surgical care, an injection site or the presence of a red flag (Cadogan et al., 2016; Lewis, 2016; Lewis et al., 2015; McClure & Michener, 2015). Conditions such as shoulder instability, calcific tendonitis, rotator cuff tendinopathy/tears, ACJ pain, referred pain from another source, glenohumeral osteoarthritis, long head of biceps tendonitis and frozen shoulder all may require different management approaches (Lewis et al., 2015). Even within these diagnoses the management can vary. For example, an acute ACJ separation would likely be managed differently to a chronic non-traumatic ACJ pain. A rotator cuff tendinopathy would have a different approach to a massive rotator cuff tear. Prognosis and treatment options can vary depending on the pathoanatomical diagnosis (McClure & Michener, 2015).

6.4 Clinical implications

This thesis has highlighted the need for clinicians to re-evaluate their practice and reflect on the validity of commonly used shoulder tests. Orthopaedic special tests are widely utilised under the assumption that they can discriminate one pathoanatomical diagnosis from another (Salamh & Lewis, 2020). However, OSTs, have been shown to load multiple tissues in the shoulder and cannot be relied on to differentiate one structure from another. The findings of this thesis challenge the traditional use of OSTs which have been over relied on in the past and oversold as 'special' diagnostic tools. In contrast, this research has

demonstrated that information gained from the wider clinical examination has a more useful role in diagnosing a condition than OSTs. Clinicians should consider all information from the clinical examination and use pattern recognition and weighting of evidence to build a differential diagnosis. The recommendation is that clinicians regard the combinations of variables that have been identified in this preliminary 10-predictor model as useful tools in the clinical reasoning process.

6.5 Recommendations for further research

The common conclusion made by the studies discussed in this thesis was that combinations of information from both the history and physical examination need to be further explored in high quality diagnostic accuracy studies (Cadogan et al., 2012; Hegedus et al., 2015; Raynor & Kuhn, 2016). Further research into clusters of clinical assessment findings such as the 10-predictor model produced here is required. This study has provided evidence that would help justify further research in this area. The study numbers were relatively small; however, the outcomes give an indication that there is value in performing a larger study to validate this combination of variables. Further validation of combinations of clinical variables to aid in the diagnosis of shoulder conditions would help to standardise clinical practice. Ultimately updated clinical guidelines to outline the most useful clinical variable combinations to facilitate accurate shoulder diagnoses are needed.

More research is required with diagnostic injections as the reference test, larger numbers of participants and rigorous methodologies.

6.6 Conclusion

This thesis has focussed on the validity of clinical tests for the diagnosis of chronic acromioclavicular joint and subacromial pain. A systematic review provided detail regarding the current evidence in respect to the reliability and diagnostic accuracy of OSTs. The reliability study was conducted, and the findings were in line with the systematic review. The conclusions were that only a small number of physical tests have sufficient reliability to warrant their use clinically. Lastly the diagnostic accuracy study demonstrated that stand-alone tests cannot influence the probability of a condition being present and recommended combining variables from the wider clinical assessment to improve their diagnostic

accuracy. This also mirrored the conclusions of the systematic review. Our diagnostic accuracy study identified a preliminary model including 10 variables from the clinical assessment that substantially improved the ability to predict a response to subacromial anaesthetic injection; and therefore confirm a subacromial diagnosis. This thesis calls for a shift away from ruling pathologies in or out on the basis of individual tests towards exploring how combining various symptoms and signs might improve our ability to identify whether a condition is present.

References

Abdelwahab, A., Ahuja, N., Iyengar, K. P., Jain, V. K., Bakti, N., & Singh, B. (2021). Traumatic rotator cuff tears - Current concepts in diagnosis and management. *Journal of Clinical Orthopaedics and Trauma, 18*, 51-55. https://doi.org/10.1016/j.jcot.2021.04.013

ACC. (2004). *The diagnosis and management of soft tissue shoulder injuries and related disorders: Best practice evidence-based guideline*. New Zealand Guidelines Group.

Ackerman, I. N., Page, R. S., Fotis, K., Schoch, P., Broughton, N., Brennan-Olsen, S. L., Bucknill, A., & Cross, E. (2018). Exploring the personal burden of shoulder pain among younger people in Australia: Protocol for a multicentre cohort study. *British Medical Journal Open, 8*(7), e021859. https://doi.org/10.1136/bmjopen-2018-021859

Ackmann, T., Schneider, K. N., Schorn, D., Rickert, C., Gosheger, G., & Liem, D. (2021, Aug). Comparison of efficacy of supraspinatus tendon tears diagnostic tests: a prospective study on the "full-can," the "empty-can," and the "Whipple" tests. *Musculoskeletal Surgery, 105*(2), 149-153. https://doi.org/10.1007/s12306-019-00631-0

Alqunaee, M., Galvin, R., & Fahey, T. (2012). Diagnostic accuracy of clinical tests for subacromial impingement syndrome: A systematic review and meta-analysis. *The Archives of Physical Medicine and Rehabilitation, 93*(2), 229-236. https://doi.org/10.1016/j.apmr.2011.08.035

Anauate Nicolao, F., Yazigi Junior, J. A., Matsunaga, F. T., Archetti Netto, N., Belloti, J. C., & Tamaoki, M. J. S. (2022). Comparing shoulder maneuvers to magnetic resonance imaging and arthroscopic findings in patients with supraspinatus tears. *World Journal of Orthopaedics, 13*(1), 102-111. https://doi.org/10.5312/wjo.v13.i1.102

Andrade, C. (2019). The P Value and Statistical Significance: Misunderstandings, Explanations, Challenges, and Alternatives. *Indian Journal of Psychological Medicine, 41*(3), 210-215. https://doi.org/10.4103/ijpsym.Ijpsym_193_19

Apeldoorn, A. T., Den Arend, M. C., Schuitemaker, R., Egmond, D., Hekman, K., Van Der Ploeg, T., Kamper, S. J., Van Tulder, M. W., & Ostelo, R. W. (2021). Interrater agreement and reliability of clinical tests for assessment of patients with shoulder pain in primary care. *Physiotherapy Theory and Practice, 37*(1), 177-196. https://doi.org/10.1080/09593985.2019.1587801

Arai, R., Nimura, A., Yamaguchi, K., Yoshimura, H., Sugaya, H., Saji, T., Matsuda, S., & Akita, K. (2014). The anatomy of the coracohumeral ligament and its relation to the subscapularis muscle. *Journal of Shoulder and Elbow Surgery, 23*(10), 1575-1581. https://doi.org/10.1016/j.jse.2014.02.009

Bak, K., Sørensen, A. K., Jørgensen, U., Nygaard, M., Krarup, A. L., Thune, C., Sloth, C., & Pedersen, S. T. (2010). The value of clinical tests in acute full-thickness tears of the supraspinatus tendon: Does a subacromial lidocaine injection help in the clinical diagnosis? A prospective study. *Arthroscopy, 26*(6), 734-742. https://doi.org/10.1016/j.arthro.2009.11.005

Balevi Batur, E., Bekin Sarıkaya, P. Z., Kaygısız, M. E., Albayrak Gezer, I., & Levendoglu, F. (2022). Diagnostic dilemma: Which clinical tests are most accurate for diagnosing supraspinatus muscle tears and tendinosis when compared to magnetic resonance imaging? *Cureus, 14*(6), e25903. https://doi.org/10.7759/cureus.25903

Barreto, R. P. G., Braman, J. P., Ludewig, P. M., Ribeiro, L. P., & Camargo, P. R. (2019). Bilateral magnetic resonance imaging findings in individuals with unilateral shoulder pain. *Journal of Shoulder and Elbow Surgery, 28*(9), 1699-1706. https://doi.org/10.1016/j.jse.2019.04.001

Baschung Pfister, P., de Bruin, E. D., Sterkele, I., Maurer, B., de Bie, R. A., & Knols, R. H. (2018). Manual muscle testing and hand-held dynamometry in people with inflammatory myopathy: An intra- and interrater reliability and validity study. *PLoS One, 13*(3), e0194531. https://doi.org/10.1371/journal.pone.0194531

Bauer, S., Okamoto, T., Babic, S. M., Coward, J. C., Coron, C., & Blakeney, W. G. (2022). Understanding shoulder pseudoparalysis: Part I: Definition to diagnosis. *EFORT open reviews, 7*(3), 214-226. https://doi.org/10.1530/eor-21-0069

Beaudreuil, J., Nizard, R., Thomas, T., Peyre, M., Liotard, J. P., Boileau, P., Marc, T., Dromard, C., Steyer, E., Bardin, T., Orcel, P., & Walch, G. (2009). Contribution of clinical tests to the diagnosis of rotator cuff disease: A systematic literature review. *Joint Bone Spine, 76*(1), 15-19. https://doi.org/10.1016/j.jbspin.2008.04.015

Bento, T., Genebra, C., Cornélio, G., Biancon, R., Simeão, S., & Vitta, A. (2019). Prevalence and factors associated with shoulder pain in the general population: A cross-sectional study. *Fisioterapia e Pesquisa, 26*, 401-406. https://doi.org/10.1590/1809-2950/18026626042019

Biederwolf, N. E. (2013). A proposed evidence-based shoulder special testing examination algorithm: Clinical utility based on a systematic review of the literature. *International Journal of Sports Physical Therapy, 8*(4), 427-440.

Blakely, T., Kvizhinadze, G., Atkinson, J., Dieleman, J., & Clarke, P. (2019, Jan). Health system costs for individual and comorbid noncommunicable diseases: An analysis of publicly funded health events from New Zealand. *PLoS Med, 16*(1), e1002716. https://doi.org/10.1371/journal.pmed.1002716

Blyth, F. M., Briggs, A. M., Schneider, C. H., Hoy, D. G., & March, L. M. (2019). The global burden of musculoskeletal pain-Where to from here? *American journal of public health, 109*(1), 35-40. https://doi.org/10.2105/AJPH.2018.304747

Boettcher, C. E., Ginn, K. A., & Cathers, I. (2009). The 'empty can' and 'full can' tests do not selectively activate supraspinatus. *Journal of Science and Medicine in Sport, 12*(4), 435-439. https://doi.org/10.1016/j.jsams.2008.09.005

Bogdanov, J., Lan, R., Chu, T. N., Bolia, I. K., Weber, A. E., & Petrigliano, F. A. (2021). Fatty degeneration of the rotator cuff: Pathogenesis, clinical implications, and future treatment. *JSES Reviews, Reports, and Techniques, 1*(4), 301-308. https://doi.org/https://doi.org/10.1016/j.xrrt.2021.05.006

Bohannon, R. W. (2018). Reliability of manual muscle testing: A systematic review. *Isokinetics and Exercise Science, 26*(4), 245-252. https://doi.org/10.3233/ies-182178

Bohannon, R. W. (2019). Considerations and practical options for measuring muscle strength: A narrative review. *BioMed Research International, 2019*, 8194537. https://doi.org/10.1155/2019/8194537

Borstad, J., & Woeste, C. (2015). The role of sensitization in musculoskeletal shoulder pain. *Brazilian Journal of Physical Therapy, 19*(4), 251-257. https://doi.org/10.1590/bjpt-rbf.2014.0100

Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., Gatsonis, C. A., Glasziou, P. P., Irwig, L., Lijmer, J. G., Moher, D., Rennie, D., de Vet, H. C. W., Kressel, H. Y., Rifai, N., Golub, R. M., Altman, D. G., Hooft, L., Korevaar, D. A., & Cohen, J. F. (2015). STARD 2015: An updated list of essential items for reporting diagnostic accuracy studies. *British Medical Journal, 351*, h5527. https://doi.org/10.1136/bmj.h5527

Bryceland, J. K., Drury, C., & Tait, G. R. (2015). Current UK practices in the management of subacromial impingement. *Shoulder & Elbow, 7*(3), 164-167. https://doi.org/10.1177/1758573215571010

Bujang, M. A., Omar, E. D., & Baharum, N. A. (2018). A review on sample size determination for Cronbach's Alpha Test: A simple guide for researchers. *The Malaysian Journal of Medical Sciences, 25*(6), 85-99. https://doi.org/10.21315/mjms2018.25.6.9

Burns, S. A., Cleland, J. A., Carpenter, K., & Mintken, P. E. (2016). Interrater reliability of the cervicothoracic and shoulder physical examination in patients with a primary complaint of shoulder pain. *Physical Therapy in Sport, 18*, 46-55. https://doi.org/10.1016/j.ptsp.2015.07.001

Byrt, T., Bishop, J., & Carlin, J. B. (1993). Bias, prevalence and kappa. *Journal of Clinical Epidemiology, 46*(5), 423-429. https://doi.org/10.1016/0895-4356(93)90018-v

Cadogan, A., Laslett, M., Hing, W., McNair, P., & Taylor, S. (2012). Clinical predictors of a positive response to guided diagnostic block into the subacromial bursa. *Journal of Rehabilitation Medicine, 44*(10), 877-884. https://doi.org/10.2340/16501977-1049

Cadogan, A., Laslett, M., Hing, W., McNair, P., & Williams, M. (2011a). Interexaminer reliability of orthopaedic special tests used in the assessment of shoulder pain. *Manual Therapy, 16*(2), 131-135. https://doi.org/10.1016/j.math.2010.07.009

Cadogan, A., Laslett, M., Hing, W., McNair, P., & Williams, M. (2011b). Reliability of a new hand-held dynamometer in measuring shoulder range of motion and strength. *Manual Therapy, 16*(1), 97-101. https://doi.org/10.1016/j.math.2010.05.005

Cadogan, A., Laslett, M., Hing, W. A., McNair, P. J., & Coates, M. H. (2011). A prospective study of shoulder pain in primary care: prevalence of imaged pathology and response to guided diagnostic blocks. *BMC Musculoskeletal Disorders, 12*, 119. https://doi.org/10.1186/1471-2474-12-119

Cadogan, A., McNair, P., Laslett, M., & Hing, W. (2013). Shoulder pain in primary care: diagnostic accuracy of clinical examination tests for non-traumatic acromioclavicular joint pain. *BMC Musculoskeletal Disorders, 14*, 156. https://doi.org/10.1186/1471-2474-14-156

Cadogan, A., McNair, P., Laslett, M., Hing, W., & Taylor, S. (2013). Diagnostic accuracy of clinical examination features for identifying large rotator cuff tears in primary health

care. *Journal of Manual & Manipulative Therapy, 21*(3), 148-159. https://doi.org/10.1179/2042618612y.0000000020

Cadogan, A., McNair, P. J., Laslett, M., & Hing, W. A. (2016). Diagnostic accuracy of clinical examination and imaging findings for identifying subacromial pain. *PLoS One, 11*(12), e0167738. https://doi.org/10.1371/journal.pone.0167738

Cadogan, C., Laslett, M., Hing, W. A., McNair, P. J., & Coates, M. H. (2011). A prospective study of shoulder pain in primary care: Prevalence of imaged pathology and response to guided diagnostic blocks. *BMC Musculoskeletal Disorders, 12*(119). https://doi.org/ 10.1186/1471-2474-12-119

Cardone, D. A., & Tallia, A. F. (2002). Joint and soft tissue injection. *American Family Physician, 66*(2), 283-288.

Chang, L. R., Anand, P., & Varacallo, M. (2022). Anatomy, Shoulder and Upper Limb, Glenohumeral Joint. In *StatPearls*. StatPearls Publishing.

Chen, G., Faris, P., Hemmelgarn, B., Walker, R. L., & Quan, H. (2009). Measuring agreement of administrative data with chart data using prevalence unadjusted and adjusted kappa. *BMC Medical Research Methodology, 9*, 5. https://doi.org/10.1186/1471-2288-9-5

Chronopoulos, E., Kim, T. K., Park, H. B., Ashenbrenner, D., & McFarland, E. G. (2004). Diagnostic value of physical tests for isolated chronic acromioclavicular lesions. *The American journal of sports medicine, 32*(3), 655-661. https://doi.org/10.1177/0363546503261723

Chung, S. W., Kim, S. H., Tae, S. K., Yoon, J. P., Choi, J. A., & Oh, J. H. (2013). Is the supraspinatus muscle atrophy truly irreversible after surgical repair of rotator cuff tears? *Clinics in Orthopedic Surgery, 5*(1), 55-65. https://doi.org/10.4055/cios.2013.5.1.55

Cibere, J., Thorne, A., Bellamy, N., Greidanus, N., Chalmers, A., Mahomed, N., Shojania, K., Kopec, J., & Esdaile, J. M. (2008). Reliability of the hip examination in osteoarthritis: effect of standardization. *Arthritis and Rheumatism, 59*(3), 373-381. https://doi.org/10.1002/art.23310

Clark, J., Sidles, J. A., & Matsen, F. A. (1990). The relationship of the glenohumeral joint capsule to the rotator cuff. *Clinial Orthopaedics & Related Research*(254), 29-34.

Clark, J. M., & Harryman, D. T., 2nd. (1992). Tendons, ligaments, and capsule of the rotator cuff. Gross and microscopic anatomy. *The Journal of Bone and Joint Surgery, 74*(5), 713-725. https://doi.org/10.2106/00004623-199274050-00010

CMDHB. (2019). *Counties Manukau Health booking & scheduling review: Ambulatory patient flow portfolio - May 2019*.

Cohen, S. P., Mao, J., Vu, T. N., Strassels, S. A., Gupta, A., Erdek, M. A., Christo, P. J., Kurihara, C., Griffith, S. R., Buckenmaier, C. C., & Chen, L. (2013). Does pain score in response to a standardized subcutaneous local anesthetic injection predict epidural steroid injection outcomes in patients with lumbosacral radiculopathy? A prospective correlational study. *Pain Medicine, 14*(3), 327-335. https://doi.org/10.1111/pme.12027

Cohen, S. P., Stojanovic, M. P., Crooks, M., Kim, P., Schmidt, R. K., Shields, C. H., Croll, S., & Hurley, R. W. (2008). Lumbar zygapophysial (facet) joint radiofrequency denervation success as a function of pain relief during diagnostic medial branch blocks: A multicenter analysis. *The Spine Journal: Official Journal of the North American Spine Society, 8*(3), 498-504. https://doi.org/10.1016/j.spinee.2007.04.022

Collin, P., Treseder, T., Denard, P. J., Neyton, L., Walch, G., & Lädermann, A. (2015). What is the best clinical test for assessment of the teres minor in massive rotator cuff tears? *Clinical Orthopaedics and Related Research, 473*(9), 2959-2966. https://doi.org/10.1007/s11999-015-4392-9

Consigliere, P., Haddo, O., Levy, O., & Sforza, G. (2018). Subacromial impingement syndrome: management challenges. *Orthopedic Research and Reviews, 10*, 83-91. https://doi.org/10.2147/orr.S157864

Cools, A. M., Cambier, D., & Witvrouw, E. E. (2008). Screening the athlete's shoulder for impingement symptoms: A clinical reasoning algorithm for early detection of shoulder pathology. *British Journal of Sports Medicine, 42*(8), 628-635. https://doi.org/10.1136/bjsm.2008.048074

Cowley, L. E., Farewell, D. M., Maguire, S., & Kemp, A. M. (2019). Methodological standards for the development and evaluation of clinical prediction rules: A review of the literature. *Diagnostic and Prognostic Research, 3*, 16. https://doi.org/10.1186/s41512-019-0060-y

Dahlhamer, J., Lucas, J., Zelaya, C., Nahin, R., Mackey, S., DeBar, L., Kerns, R., Von Korff, M., Porter, L., & Helmick, C. (2018). Prevalence of chronic pain and high-impact chronic

pain among adults - United States, 2016. *Morbidity and Mortality Weekly Report, 67*(36), 1001-1006. https://doi.org/10.15585/mmwr.mm6736a2

Dakkak, A., Krill, M. K., Krill, M. L., Nwachukwu, B., & McCormick, F. (2021). Evidence-based physical examination for the diagnosis of subscapularis tears: A systematic review. *Sports Health, 13*(1), 78-84. https://doi.org/10.1177/1941738120936232

Daniels, E. W., Cole, D., Jacobs, B., & Phillips, S. F. (2018). Existing evidence on ultrasound-guided injections in sports medicine. *Orthopaedic Journal of Sports Medicine, 6*(2), 2325967118756576. https://doi.org/10.1177/2325967118756576

Datta, S., Lee, M., Falco, F. J., Bryce, D. A., & Hayek, S. (2009). Systematic assessment of diagnostic accuracy and therapeutic utility of lumbar facet joint interventions. *Pain Physician, 12*(2), 437–460.

Davison, S. N., Rathwell, S., Ghosh, S., George, C., Pfister, T., & Dennett, L. (2021). The Prevalence and Severity of Chronic Pain in Patients With Chronic Kidney Disease: A Systematic Review and Meta-Analysis. *Canadian Journal of Kidney Health and Disease, 8*, 2054358121993995. https://doi.org/10.1177/2054358121993995

Dawson, J., Rogers, K., Fitzpatrick, R., & Carr, A. (2009). The Oxford shoulder score revisited. *Archives of Orthopaedic and Trauma Surgery, 129*(1), 119-123. https://doi.org/10.1007/s00402-007-0549-7

De Vries, H., Elliott, M., Kanouse, D., & Teleki, S. (2008). Using pooled kappa to summarize interrater agreement across many items. *Field Methods 20*, 272-282. https://doi.org/10.1177/1525822X08317166

de Winter, A. F., Jans, M. P., Scholten, R. J., Devillé, W., van Schaardenburg, D., & Bouter, L. M. (1999). Diagnostic classification of shoulder disorders: Interobserver agreement and determinants of disagreement. *Annals of the Rheumatic Diseases, 58*(5), 272-277. https://doi.org/10.1136/ard.58.5.272

De Yang Tien, J., & Tan, A. H. C. (2014). Shoulder impingement syndrome, a common affliction of the shoulder: A comprehensive review. *Proceedings of Singapore Healthcare, 23*(4), 297-305. https://doi.org/10.1177/201010581402300406

Denegar, C. R., & Fraser, M. (2006). How useful are physical examination procedures? Understanding and applying likelihood ratios. *Journal of Athletic Training, 41*(2), 201-206.

Derby, R., Melnik, I., Lee, J.-E., & Lee, S.-H. (2012). Correlation of lumbar medial branch neurotomy results with diagnostic medial branch block cutoff values to optimize therapeutic outcome. *Pain Medicine, 13*(12), 1533-1546. https://doi.org/10.1111/j.1526-4637.2012.01500.x

Deyo, R. A. (2002). Cascade effects of medical technology. *Annual Review of Public Health, 23*, 23-44. https://doi.org/10.1146/annurev.publhealth.23.092101.134534

Dinnes, J., Loveman, E., McIntyre, L., & Waugh, N. (2003). The effectiveness of diagnostic tests for the assessment of shoulder pain due to soft tissue disorders: A systematic review. *Health Technology Assessment, 7*(29), iii-166. https://doi.org/10.3310/hta7290

Djade, C. D., Porgo, T. V., Zomahoun, H. T. V., Perrault-Sullivan, G., & Dionne, C. E. (2020). Incidence of shoulder pain in 40 years old and over and associated factors: A systematic review. *European Journal of Pain, 24*(1), 39-50. https://doi.org/10.1002/ejp.1482

Doxey, R., Thiese, M. S., & Hegmann, K. T. (2018). Reliability of common provocative tests for shoulder tendinitis. *The Journal of Occupational and Environmental Medicine, 60*(12), 1063-1066. https://doi.org/10.1097/jom.0000000000001430

Dutil, É., Bottari, C., & Auger, C. (2017). Test-retest reliability of a measure of independence in everyday activities: The ADL profile. *Occupational Therapy International, 2017*, 3014579. https://doi.org/10.1155/2017/3014579

Edwards, P., Ebert, J., Joss, B., Bhabra, G., Ackland, T., & Wang, A. (2016). Exercise rehabilitation in the non-operative management of rotator cuff tears: A review of the literature. *International Journal of Sports Physical Therapy,, 11*(2), 279-301.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological review, 100*(3), 363-406. https://doi.org/https://doi.org/10.1037/0033-295X.100.3.363

Farrell, G., Watson, L., & Devan, H. (2019). Current evidence for nonpharmacological interventions and criteria for surgical management of persistent acromioclavicular joint osteoarthritis: A systematic review. *Shoulder & Elbow, 11*(6), 395-410. https://doi.org/10.1177/1758573219840673

Fayaz, A., Croft, P., Langford, R. M., Donaldson, L. J., & Jones, G. T. (2016). Prevalence of chronic pain in the UK: A systematic review and meta-analysis of population studies. *British Medical Journal Open, 6*(6), e010364. https://doi.org/10.1136/bmjopen-2015-010364

Ferenczi, A., Ostertag, A., Lasbleiz, S., Petrover, D., Yelnik, A., Richette, P., Bardin, T., Orcel, P., & Beaudreuil, J. (2018). Reproducibility of sub-acromial impingement tests, including a new clinical manoeuver. *Annals of Physical and Rehabilitation Medicine, 61*(3), 151-155. https://doi.org/10.1016/j.rehab.2018.01.005

Fine, P. G. (2011). Long-term consequences of chronic pain: Mounting evidence for pain as a neurological disease and parallels with other chronic disease states. *Pain Medicine, 12*(7), 996-1004. https://doi.org/10.1111/j.1526-4637.2011.01187.x

Finniss, D., Nicholas, M., Brooker, C., Cousins, M., & Benedetti, F. (2019). Magnitude, response, and psychological determinants of placebo effects in chronic low-back pain: a randomised, double-blinded, controlled trial. *Pain Reports, 4*(3). https://doi.org/10.1097/PR9.0000000000000744

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika, 80*(1), 27-38. https://doi.org/https://doi.org/10.2307/2336755

Frantz, T., Ramkumar, P. N., Frangiamore, S., Jones, G., Soloff, L., Kvit, A., Curriero, F. C., & Schickendantz, M. (2021). Epidemiology of acromioclavicular joint injuries in professional baseball: Analysis from the major league baseball health and injury tracking system. *Journal of Shoulder and Elbow Surgery, 30*(1), 127-133. https://doi.org/10.1016/j.jse.2020.04.029

Friedman, D. J., & Khan, K. M. (2019). Preventing overdiagnosis and the harms of too much sport and exercise medicine. *British Journal of Sports Medicine, 53*(20), 1314-1318. https://doi.org/10.1136/bjsports-2018-100039

Fritz, J. M., & Wainner, R. S. (2001). Examining diagnostic tests: An evidence-based perspective. *Physical Therapy in Sport, 81*(9), 1546-1564. https://doi.org/10.1093/ptj/81.9.1546

Garner, P., Hopewell, S., Chandler, J., MacLehose, H., Schünemann, H. J., Akl, E. A., Beyene, J., Chang, S., Churchill, R., Dearness, K., Guyatt, G., Lefebvre, C., Liles, B., Marshall, R., Martínez García, L., Mavergames, C., Nasser, M., Qaseem, A., Sampson, M., ..., & Schünemann, H. J. (2016). When and how to update systematic reviews: Consensus and checklist. *British Medical Journal (Clinical research ed.), 354*, i3507. https://doi.org/10.1136/bmj.i3507

Gerber, C., Galantay, R. V., & Hersche, O. (1998). The pattern of pain produced by irritation of the acromioclavicular joint and the subacromial space. *Journal of Shoulder and Elbow Surgery, 7*(4), 352-355. https://doi.org/10.1016/s1058-2746(98)90022-2

Gismervik, S., Drogset, J. O., Granviken, F., Rø, M., & Leivseth, G. (2017). Physical examination tests of the shoulder: A systematic review and meta-analysis of diagnostic test performance. *BMC Musculoskeletal Disorders, 18*(1), 41. https://doi.org/10.1186/s12891-017-1400-0

Godenèche, A., Nové-Josserand, L., Audebert, S., Toussaint, B., Denard, P. J., Arthroscopy, F. S. f., & Lädermann, A. (2017). Relationship between subscapularis tears and injuries to the biceps pulley. *Knee Surgery, Sports Traumatology, Arthroscopy, 25*, 2114-2120. https://doi.org/10.1007/s00167-016-4374-9

Graham, R., & Masters-Awatere, B. (2020). Experiences of Māori of Aotearoa New Zealand's public health system: A systematic review of two decades of published qualitative research. *Australian and New Zealand Journal of Public Health, 44*. https://doi.org/10.1111/1753-6405.12971

Green, R., Shanley, K., Taylor, N. F., & Perrott, M. (2008). The anatomical basis for clinical tests assessing musculoskeletal function of the shoulder. *Physical Therapy Reviews, 13*(1), 17-24. https://doi.org/10.1179/174328808X251966

Guerchet, M., Mayston, R., & Prina, A. M. (2020). Conducting research in low-and middle-income settings. *Oxford Textbook of Neurologic and Neuropsychiatric Epidemiology*, 47. https://doi.org/10.2471/BLT.06.032375

Hackshaw, A. (2008). Small studies: Strengths and limitations. *European Respiratory Journal, 32*(5), 1141-1143. https://doi.org/10.1183/09031936.00136408

Hanchard, N. C., Goodchild, L., Thompson, J., O'brien, T., Davison, D., & Richardson, C. (2012). Evidence-based clinical guidelines for the diagnosis, assessment and physiotherapy management of contracted (frozen) shoulder: Quick reference summary. *Physiotherapy, 98*(2), 117-120. https://doi.org/10.1016/j.physio.2012.01.001

Harrell, F. E. (2015). *Regression modeling strategies with applications to linear models, logistic and ordinal regression, and survival analysis* (2nd ed.). Spinger. https://doi.org/10.1007/978-3-319-19425-7.

Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCap): A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics, 42*(2), 377-381. https://doi.org/10.1016/j.jbi.2008.08.010

Hayes, K., Walton, J. R., Szomor, Z. L., & Murrell, G. A. (2002). Reliability of 3 methods for assessing shoulder strength. *Journal of Shoulder and Elbow Surgery, 11*(1), 33-39. https://doi.org/10.1067/mse.2002.119852

Hegedus, E. J., Cook, C., Lewis, J., Wright, A., & Park, J. Y. (2015). Combining orthopedic special tests to improve diagnosis of shoulder pathology. *Physical Therapy in Sport, 16*(2), 87-92. https://doi.org/10.1016/j.ptsp.2014.08.001

Hegedus, E. J., Goode, A., Campbell, S., Morin, A., Tamaddoni, M., Moorman, C. T., & Cook, C. (2008). Physical examination tests of the shoulder: A systematic review with meta-analysis of individual tests. *British Journal of Sports Medicine, 42*(2), 80-92 https://doi.org/10.1136/bjsm.2007.038406

Hegedus, E. J., Goode, A. P., Cook, C. E., Michener, L., Myer, C. A., Myer, D. M., & Wright, A. A. (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. *British Journal of Sports Medicine, 46*(14), 964-978. https://doi.org/10.1136/bjsports-2012-091066

Hegedus, E. J., Wright, A. A., & Cook, C. (2017). Orthopaedic special tests and diagnostic accuracy studies: House wine served in very cheap containers. *British Journal of Sports Medicine, 51*(22), 1578-1579. https://doi.org/10.1136/bjsports-2017-097633

Heidar Abady, A., Rosedale, R., Chesworth, B. M., Rotondi, M. A., & Overend, T. J. (2018). Consistency of commonly used orthopedic special tests of the shoulder when used with the McKenzie system of mechanical diagnosis and therapy. *Musculoskeletal Science & Practice, 33*, 11-17. https://doi.org/10.1016/j.msksp.2017.10.001

Helbach, J., Pieper, D., Mathes, T., Rombey, T., Zeeb, H., Allers, K., & Hoffmann, F. (2022). Restrictions and their reporting in systematic reviews of effectiveness: An observational study. *BMC Medical Research Methodology, 22*(1), 230. https://doi.org/10.1186/s12874-022-01710-w

Hermans, J., Luime, J. J., Meuffels, D. E., Reijman, M., Simel, D. L., & Bierma-Zeinstra, S. M. (2013). Does this patient with shoulder pain have rotator cuff disease? The rational clinical examination systematic review. *Journal of the American Medical Association, 310*(8), 837-847. https://doi.org/10.1001/jama.2013.276187

Hespanhol, L., Vallio, C. S., Costa, L. M., & Saragiotto, B. T. (2019). Understanding and interpreting confidence and credible intervals around effect estimates. *Brazilian Journal of Physical Therapy, 23*(4), 290-301. https://doi.org/10.1016/j.bjpt.2018.12.006

Hodgson, R. J., O'Connor, P. J., Hensor, E. M., Barron, D., & Robinson, P. (2012). Contrast-enhanced MRI of the subdeltoid, subacromial bursa in painful and painless rotator cuff tears. *The British Journal of Radiology, 85*(1019), 1482-1487. https://doi.org/10.1259/bjr/45423226

Holtby, R., & Razmjou, H. (2004). Validity of the supraspinatus test as a single clinical test in diagnosing patients with rotator cuff pathology. *The Journal of Orthopaedic and Sports Physical Therapy, 34*(4), 194-200. https://doi.org/10.2519/jospt.2004.34.4.194

Hoo, Z. H., Candlish, J., & Teare, D. (2017). What is an ROC curve? *Emergency Medicine Journal: EMJ, 34*(6), 357-359. https://doi.org/10.1136/emermed-2017-206735

Hughes, P. C., Taylor, N. F., & Green, R. A. (2008). Most clinical tests cannot accurately diagnose rotator cuff pathology: A systematic review. *Australian Journal of Physiotherapy 54*(3), 159-170. https://doi.org/10.1016/s0004-9514(08)70022-9

Innocenti, T., Ristori, D., Miele, S., & Testa, M. (2019). The management of shoulder impingement and related disorders: A systematic review on diagnostic accuracy of physical tests and manual therapy efficacy. *Journal of Bodywork and Movement Therapies, 23*(3), 604-618. https://doi.org/10.1016/j.jbmt.2018.08.002

Institute for Health Metrics and Evaluation. (2019). *Musculoskeletal disorders — Level 2 cause*. Retrieved 22 August 2021 from http://www.healthdata.org/results/gbd_summaries/2019/musculoskeletal-disorders-level-2-cause

Ioannidis, J. P. (2007). Limitations are not properly acknowledged in the scientific literature. *Journal of Clinical Epidemiology, 60*(4), 324-329. https://doi.org/10.1016/j.jclinepi.2006.09.011

Itoi, E., Kido, T., Sano, A., Urayama, M., & Sato, K. (1999). Which is more useful, the "full can test" or the "empty can test," in detecting the torn supraspinatus tendon? *The American Journal of Sports Medicine, 27*(1), 65-68. https://doi.org/10.1177/03635465990270011901

Itoi, E., Minagawa, H., Yamamoto, N., Seki, N., & Abe, H. (2006). Are pain location and physical examinations useful in locating a tear site of the rotator cuff? *The American Journal of Sports Medicine, 34*(2), 256-264. https://doi.org/10.1177/0363546505280430

Jaeschke, R., Guyatt, G. H., & Sackett, D. L. (1994). Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. *Journal of the American Medical Association, 271*(9), 703-707. https://doi.org/10.1001/jama.271.9.703

Jain, N. B., Luz, J., Higgins, L. D., Dong, Y., Warner, J. J., Matzkin, E., & Katz, J. N. (2017). The diagnostic accuracy of special tests for rotator cuff tear: The ROW cohort study. *American Journal of Physical Medicine & Rehabilitation, 96*(3), 176-183. https://doi.org/10.1097/PHM.0000000000000566

James, M., Stokes, E. A., Thomas, E., Dziedzic, K., & Hay, E. M. (2005). A cost consequences analysis of local corticosteroid injection and physiotherapy for the treatment of new episodes of unilateral shoulder pain in primary care. *Rheumatology (Oxford), 44*(11), 1447-1451. https://doi.org/10.1093/rheumatology/kei043

Jancuska, J., Matthews, J., Miller, T., Kluczynski, M. A., & Bisson, L. J. (2018, Sep). A Systematic Summary of Systematic Reviews on the Topic of the Rotator Cuff. *Orthopaedic Journal of Sports Medicine, 6*(9), 2325967118797891. https://doi.org/10.1177/2325967118797891

Javed, S., Sadozai, Z., Javed, A., Din, A., & Schmitgen, G. (2017). Should all acromioclavicular joint injections be performed under image guidance? *Journal of Orthopaedic Surgery, 25*(3), 2309499017731633. https://doi.org/10.1177/2309499017731633

Johansson, K., & Ivarson, S. (2009). Intra- and interexaminer reliability of four manual shoulder maneuvers used to identify subacromial pain. *Manual Therapy, 14*(2), 231-239. https://doi.org/10.1016/j.math.2008.03.003

Kane, D., & Koski, J. (2016). Musculoskeletal interventional procedures: With or without imaging guidance? *Best Practice & Research: Clinical Rheumatology, 30*(4), 736-750. https://doi.org/10.1016/j.berh.2016.09.012

Kaul, A., Mandal, S., Davidov, O., & Peddada, S. D. (2017). Analysis of microbiome data in the presence of excess zeros. *Frontiers in microbiology, 8*, 2114. https://doi.org/10.3389/fmicb.2017.02114

Kennedy, C. A., Manno, M., Hogg-Johnson, S., Haines, T., Hurley, L., McKenzie, D., & Beaton, D. E. (2006). Prognosis in soft tissue disorders of the shoulder: Predicting both change in disability and level of disability after treatment. *Physical Therapy, 86*(7), 1013-1032; discussion 1033-1017. https://doi.org/10.1093/ptj/86.7.1013

Klatte-Schulz, F., Thiele, K., Scheibel, M., Duda, G. N., & Wildemann, B. (2022). Subacromial Bursa: A Neglected Tissue Is Gaining More and More Attention in Clinical and Experimental Research. *Cells, 11*(4). https://doi.org/10.3390/cells11040663

Klintberg, I. H., Cools, A. M., Holmgren, T. M., Holzhausen, A. C., Johansson, K., Maenhout, A. G., Moser, J. S., Spunton, V., & Ginn, K. (2015). Consensus for physiotherapy for shoulder pain. *International Orthopaedics, 39*(4), 715-720. https://doi.org/10.1007/s00264-014-2639-9

Krill, M. K., Rosas, S., Kwon, K., Dakkak, A., Nwachukwu, B. U., & McCormick, F. (2018). A concise evidence-based physical examination for diagnosis of acromioclavicular joint pathology: A systematic review. *The Physician and Sports Medicine, 46*(1), 98-104. https://doi.org/10.1080/00913847.2018.1413920

Lädermann, A., Collin, P., Zbinden, O., Meynard, T., Saffarini, M., & Chiu, J. C. (2021). Diagnostic accuracy of clinical tests for subscapularis tears: A systematic review and meta-analysis. *Orthopaedic Journal of Sports Medicine, 9*(9), 23259671211042011. https://doi.org/10.1177/23259671211042011

Lädermann, A., Meynard, T., Denard, P. J., Ibrahim, M., Saffarini, M., & Collin, P. (2021). Reliable diagnosis of posterosuperior rotator cuff tears requires a combination of clinical tests. *Knee Surgery, Sports Traumatology, Arthroscopy, 29*(7), 2118-2133. https://doi.org/10.1007/s00167-020-06136-9

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159-174. https://doi.org/10.2307/2529310

Lange, T., Matthijs, O., Jain, N. B., Schmitt, J., Lutzner, J., & Kopkow, C. (2017). Reliability of specific physical examination tests for the diagnosis of shoulder pathologies: A systematic review and meta-analysis. *British Journal of Sports Medicine, 51*(6), 511-518. https://doi.org/10.1136/bjsports-2016-096558

Laron, D., Samagh, S. P., Liu, X., Kim, H. T., & Feeley, B. T. (2012). Muscle degeneration in rotator cuff tears. *Journal of Shoulder and Elbow Surgery, 21*(2), 164-174. https://doi.org/10.1016/j.jse.2011.09.027

Latremoliere, A., & Woolf, C. J. (2009). Central sensitization: A generator of pain hypersensitivity by central neural plasticity. *The Journal of Pain, 10*(9), 895-926. https://doi.org/10.1016/j.jpain.2009.06.012

Lee, S. H., Choi, H. H., & Lee, D. G. (2020). Effectiveness of new nerve blocks method on the articular branches of the suprascapular and subscapular nerves to treat shoulder pain. *Medicine, 99*(35). https://doi.org/10.1097/MD.0000000000022050

Lewis, J. (2015). Frozen shoulder contracture syndrome - Aetiology, diagnosis and management. *Manual Therapy, 20*(1), 2-9. https://doi.org/10.1016/j.math.2014.07.006

Lewis, J. (2016). Rotator cuff related shoulder pain: Assessment, management and uncertainties. *Manual Therapy, 23*, 57-68. https://doi.org/10.1016/j.math.2016.03.009

Lewis, J., McCreesh, K., Roy, J. S., & Ginn, K. (2015). Rotator cuff tendinopathy: Navigating the diagnosis-management conundrum. *The Journal of Orthopaedic and Sports Physical Therapy, 45*(11), 923-937. https://doi.org/10.2519/jospt.2015.5941

Lewis, J. S. (2009). Rotator cuff tendinopathy/subacromial impingement syndrome: Is it time for a new method of assessment? *British Journal of Sports Medicine, 43*(4), 259-264. https://doi.org/10.1136/bjsm.2008.052183

Li, J. L., Ge, L., Ma, J. C., Zeng, Q. L., Yao, L., An, N., Ding, J. X., Gan, Y. H., & Tian, J. H. (2014). Quality of reporting of systematic reviews published in "evidence-based" Chinese journals. *Systematic Reviews, 3*, 58. https://doi.org/10.1186/2046-4053-3-58

Lucas, N., Macaskill, P., Irwig, L., Moran, R., Rickards, L., Turner, R., & Bogduk, N. (2013). The reliability of a quality appraisal tool for studies of diagnostic reliability (QAREL). *BMC Medical Research Methodology, 13*(1), 111. https://doi.org/10.1186/1471-2288-13-111

Malfair, D. (2008). Therapeutic and diagnostic joint injections. *Radiologic Clinics of North America, 46*(3), 439-453, v. https://doi.org/10.1016/j.rcl.2008.02.007

Manchikanti, L., Kaye, A. D., Soin, A., Albers, S. L., & Beall, D. (2020). Comprehensive evidence-based guidelines for facet joint interventions in the management of chronic spinal pain: American Society of Interventional Pain Physicians (ASIPP) guidelines. *Pain Physician, 23*(5).

Mandrekar, J. N. (2010). Receiver operating characteristic curve in diagnostic test assessment. *Journal of Thoracic Oncology, 5*(9), 1315-1316. https://doi.org/10.1097/JTO.0b013e3181ec173d

Mani, S., Sharma, S., Omar, B., Paungmali, A., & Joseph, L. (2017). Validity and reliability of Internet-based physiotherapy assessment for musculoskeletal disorders: A systematic review. *Journal of Telemedicine and Telecare, 23*(3), 379-391. https://doi.org/10.1177/1357633x16642369

Marks, D., Thomas, M., Newans, T., & Bisset, L. (2023). Immediate response to injection is associated with conservative care outcomes at 12 weeks in subacromial shoulder pain. *Musculoskeletal Science & Practice, 64*, 102726. https://doi.org/10.1016/j.msksp.2023.102726

May, S., Chance-Larsen, K., Littlewood, C., Lomas, D., & Saad, M. (2010). Reliability of physical examination tests used in the assessment of patients with shoulder problems: A systematic review. *Physiotherapy, 96*(3), 179-190. https://doi.org/10.1016/j.physio.2009.12.002

McClure, P. W., & Michener, L. A. (2015). Staged approach for rehabilitation classification: Shoulder disorders (STAR-Shoulder). *Physical Therapy, 95*(5), 791-800. https://doi.org/10.2522/ptj.20140156

McFarland, E., Bernard, J., Dein, E., & Johnson, A. (2017). Diagnostic injections about the shoulder. *The Journal of the American Academy of Orthopaedic Surgeons, 25*(12), 799-807. https://doi.org/10.5435/jaaos-d-16-00076

McGee, S. (2002). Simplifying likelihood ratios. *Journal of General Internal Medicine, 17*(8), 646-649. https://doi.org/10.1046/j.1525-1497.2002.10750.x

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica (Zagreb), 22*(3), 276-282. https://doi.org/10.11613/BM.2012.031

Medsafe. (2018). *New Zealand Data Sheet: Xylocaine*. Retrieved February 11, 2023, from https://www.medsafe.govt.nz/profs/datasheet/x/XylocaineAndAdrenalineinj.pdf

Michener, L. A., Walsworth, M. K., Doukas, W. C., & Murphy, K. P. (2009). Reliability and diagnostic accuracy of 5 physical examination tests and combination of tests for subacromial impingement. *The Archives of Physical Medicine and Rehabilitation, 90*(11), 1898-1903. https://doi.org/10.1016/j.apmr.2009.05.015

Milgrom, C., Schaffler, M., Gilbert, S., & van Holsbeeck, M. (1995). Rotator-cuff changes in asymptomatic adults. The effect of age, hand dominance and gender. *The Journal of Bone and Joint Surgery, 77*(2), 296-298. https://doi.org/10.1302/0301-620X.77B2.7706351

Ministry of Health. (2016). *Health Loss in New Zealand 1990–2013: A report from the New Zealand Burden of Diseases, Injuries and Risk Factors Study*. Ministry of Health. Retrieved 4 April 2021 from https://www.health.govt.nz/publication/health-loss-new-zealand-1990-2013

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine, 6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

Nahm, F. S. (2022). Receiver operating characteristic curve: Overview and practical use for clinicians. *Korean Journal of Anesthesiology, 75*(1), 25-36. https://doi.org/10.4097/kja.21209

Newcombe, R. (1992, 10/01). Practical Statistics for Medical Research. *Journal of Epidemiology & Community Health, 46*. https://doi.org/10.1136/jech.46.5.549-a

Nomden, J. G., Slagers, A. J., Bergman, G. J., Winters, J. C., Kropmans, T. J., & Dijkstra, P. U. (2009). Interobserver reliability of physical examination of shoulder girdle. *Manual Therapy, 14*(2), 152-159. https://doi.org/10.1016/j.math.2008.01.005

Nurjannah, I., & Siwi, S. M. (2017). Guidelines for analysis on measuring interrater reliability of nursing outcome classification. *International Journal of Research in Medical Sciences, 5*(4), 1169-1175. https://doi.org/10.18203/2320-6012.ijrms20171220

O'Brien, S. J., Pagnani, M. J., Fealy, S., McGlynn, S. R., & Wilson, J. B. (1998). The active compression test: A new and effective test for diagnosing labral tears and acromioclavicular joint abnormality. *American Journal of Sports Medicine, 26*(5), 610-613. https://doi.org/10.1177/03635465980260050201

O'Holleran, J. D., Kocher, M. S., Horan, M. P., Briggs, K. K., & Hawkins, R. J. (2005). Determinants of patient satisfaction with outcome after rotator cuff surgery. *Journal of Bone and Joint Surgery. American volume, 87*(1), 121-126. https://doi.org/10.2106/jbjs.C.01316

Ostor, A. J., Richards, C. A., Prevost, A. T., Hazleman, B. L., & Speed, C. A. (2004). Interrater reproducibility of clinical tests for rotator cuff lesions. *Annals of the Rheumatic Diseases, 63*(10), 1288-1292. https://doi.org/10.1136/ard.2003.014712

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ..., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *International Journal of Surgery, 88*, 105906. https://doi.org/https://doi.org/10.1016/j.ijsu.2021.105906

Papadonikolakis, A., McKenna, M., Warme, W., Martin, B. I., & Matsen, F. A. (2011). Published evidence relevant to the diagnosis of impingement syndrome of the shoulder. *The Journal of Bone and Joint Surgery. American volume, 93*(19), 1827-1832. https://doi.org/10.2106/jbjs.J.01748

Parikh, R., Parikh, S., Arun, E., & Thomas, R. (2009). Likelihood ratios: Clinical application in day-to-day practice. *Indian Journal of Ophthalmology, 57*(3), 217-221. https://doi.org/10.4103/0301-4738.49397

Park, H. B., Yokota, A., Gill, H. S., El Rassi, G., & McFarland, E. G. (2005). Diagnostic accuracy of clinical tests for the different degrees of subacromial impingement syndrome. *Journal of Bone and Joint Surgery, 87*(7), 1446-1455. https://doi.org/10.2106/JBJS.D.02335

Park, J. Y., Chung, K. T., & Yoo, M. J. (2004). A serial comparison of arthroscopic repairs for partial- and full-thickness rotator cuff tears. *Arthroscopy, 20*(7), 705-711. https://doi.org/10.1016/j.arthro.2004.06.013

Perme, M. P., & Manevski, D. (2019). Confidence intervals for the Mann-Whitney test. *Statistical Methods in Medical Research, 28*(12), 3755-3768. https://doi.org/10.1177/0962280218814556

Powell, J. W., & Huijbregts, P. A. (2006). Concurrent criterion-related validity of acromioclavicular joint physical examination tests: A systematic review. *Journal of Manual & Manipulative Therapy, 14*(2), 19E-29E. https://doi.org/10.1179/jmt.2006.14.2.19E

Power, M., Fell, G., & Wright, M. (2013). Principles for high-quality, high-value testing. *British Medical Journal, 18*(1), 5-10. https://doi.org/http://dx.doi.org/10.1136/eb-2012-100645

R Core Team. (2020). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. https://www.R-project.org/

Raynor, M. B., & Kuhn, J. E. (2016). Utility of features of the patient's history in the diagnosis of atraumatic shoulder pain: A systematic review. *Journal of Shoulder and Elbow Surgery, 25*(4), 688-694. https://doi.org/10.1016/j.jse.2015.09.023

Razmjou, H., Holtby, R., & Myhr, T. (2004). Pain provocative shoulder tests: Reliability and validity of the impingement tests. *Physiotherapy Canada, 56*. https://doi.org/10.2310/6640.2004.00027

Revelle, W. (2013). psych: Procedures for Psychological, psychometric, and personality research. R Package Version 1.0–95. *Evanston, Illinois*.

Roache, P. B. (2021). Anterior cable tears in arthroscopic rotator cuff repairs. *Arthroscopy, Sports Medicine, and Rehabilitation, 3*(3), e695-e705. https://doi.org/https://doi.org/10.1016/j.asmr.2021.01.007

Safari, S., Baratloo, A., Elfil, M., & Negida, A. (2015). Evidence based emergency medicine part 2: Positive and negative predictive values of diagnostic tests. *Emergency 3*(3), 87-88. https://doi.org/10.22037/aaem.v3i3.274

Safiri, S., Kolahi, A. A., Cross, M., Hill, C., Smith, E., Carson-Chahhoud, K., Mansournia, M. A., Almasi-Hashiani, A., Ashrafi-Asgarabad, A., Kaufman, J., Sepidarkish, M., Shakouri, S. K., Hoy, D., Woolf, A. D., March, L., Collins, G., & Buchbinder, R. (2021). Prevalence, deaths, and disability-adjusted life years due to musculoskeletal disorders for 195 countries and territories 1990-2017. *Arthritis & Rheumatology, 73*(4), 702-714. https://doi.org/10.1002/art.41571

Sahu, D., & Phadnis, A. (2021). Revisiting the rotator cuff footprint. *Journal of Clinical Orthopaedics and Trauma, 21*, 101514. https://doi.org/10.1016/j.jcot.2021.101514

Salaffi, F., Ciapetti, A., Carotti, M., Gasparini, S., Filippucci, E., & Grassi, W. (2010). Clinical value of single versus composite provocative clinical tests in the assessment of painful shoulder. *Journal of Clinical Rheumatology, 16*(3), 105-108. https://doi.org/10.1097/RHU.0b013e3181cf8392

Salamh, P., & Lewis, J. (2020). It is time to put special tests for rotator cuff-related shoulder pain out to pasture. *The Journal of Orthopaedic and Sports Physical Therapy, 50*(5), 222-225. https://doi.org/10.2519/jospt.2020.0606

Scheel, C., Mecham, J., Zuccarello, V., & Mattes, R. (2018). An evaluation of the inter-rater and intra-rater reliability of OccuPro's functional capacity evaluation. *Work, 60*(3), 465-473. https://doi.org/10.3233/wor-182754

Schervish, M. J. (1996). P values: What they are and what they are not. *The American Statistician, 50*(3), 203-206. https://doi.org/10.1080/00031305.1996.10474380

Schmidt, M., Enger, M., Pripp, A. H., Nordsletten, L., Moosmayer, S., Melhuus, K., & Brox, J. I. (2021). Interrater reliability of physical examination tests in the acute phase of shoulder injuries. *BMC Musculoskeletal Disorders, 22*(1), 1-9. https://doi.org/10.1186/s12891-021-04659-x.

Schneider, B. J., Huynh, L., Levin, J., Rinkaekan, P., Kordi, R., & Kennedy, D. J. (2018). Does immediate pain relief after an injection into the sacroiliac joint with anesthetic and corticosteroid predict subsequent pain relief? *Pain Medicine, 19*(2), 244-251. https://doi.org/10.1093/pm/pnx104

Seo, S. S., Choi, J. S., An, K. C., Kim, J. H., & Kim, S. B. (2012). The factors affecting stiffness occurring with rotator cuff tear. *Journal of Shoulder and Elbow Surgery, 21*(3), 304-309. https://doi.org/10.1016/j.jse.2011.04.011

Shaffer, B. S. (1999). Painful conditions of the acromioclavicular joint. *Journal of the American Academy of Orthopaedic Surgeons, 7*(3), 176-188. https://doi.org/10.5435/00124635-199905000-00004

Sharma, S. P., Bærheim, A., & Kvåle, A. (2015). Passive range of motion in patients with adhesive shoulder capsulitis, an intertester reliability study over eight weeks. *BMC Musculoskeletal Disorders, 16*, 37. https://doi.org/10.1186/s12891-015-0495-4

Sher, J. S., Uribe, J. W., Posada, A., Murphy, B. J., & Zlatkin, M. B. (1995). Abnormal findings on magnetic resonance images of asymptomatic shoulders. *The Journal of Bone and Joint Surgery. American volume, 77*(1), 10-15. https://doi.org/10.2106/00004623-199501000-00002

Shreffler, J., & Huecker, M. R. (2023). *Diagnostic testing accuracy: Sensitivity, specificity, predictive values and likelihood ratios*. StatPearls Publishing. Retrieved 13 May 2023 from https://www.ncbi.nlm.nih.gov/books/NBK557491/

Shubin Stein, B. E., Wiater, J. M., Pfaff, H. C., Bigliani, L. U., & Levine, W. N. (2001). Detection of acromioclavicular joint pathology in asymptomatic shoulders with magnetic

resonance imaging. *Journal of Shoulder and Elbow Surgery, 10*(3), 204-208. https://doi.org/10.1067/mse.2001.113498

Šimundić, A.-M. (2009). Measures of Diagnostic Accuracy: Basic Definitions. *Electronic Journal of the International Federation of Clinical Chemistry and Laboratory Medicine, 19*(4), 203-211. https://pubmed.ncbi.nlm.nih.gov/27683318

Sirin, E., Aydin, N., & Mert Topkar, O. (2018). Acromioclavicular joint injuries: Diagnosis, classification and ligamentoplasty procedures. *EFORT open reviews, 3*(7), 426-433. https://doi.org/10.1302/2058-5241.3.170027

Skedros, J. G., & Pitts, T. C. (2007). Temporal variations in a modified Neer impingement test can confound clinical interpretation. *Clinical Orthopaedics and Related Research 460*, 130-136. https://doi.org/10.1097/BLO.0b013e31804a8581

Skinner, D., Skinner, D. V., Swain, A., Robertson, C., & Peyton, J. R. (1997). *Cambridge textbook of accident and emergency medicine*. Cambridge University Press.

Sørensen, A. K., Bak, K., Krarup, A. L., Thune, C. H., Nygaard, M., Jørgensen, U., Sloth, C., & Torp-Pedersen, S. (2007). Acute rotator cuff tear: Do we miss the early diagnosis? A prospective study showing a high incidence of rotator cuff tears after shoulder trauma. *Journal of Shoulder and Elbow Surgery, 16*(2), 174-180. https://doi.org/10.1016/j.jse.2006.06.010

Strobel, K., Pfirrmann, C. W., Zanetti, M., Nagy, L., & Hodler, J. (2003). MRI features of the acromioclavicular joint that predict pain relief from intraarticular injection. *The American Journal of Roentgenology, 181*(3), 755-760. https://doi.org/10.2214/ajr.181.3.1810755

Struyf, F., Geraets, J., Noten, S., Meeus, M., & Nijs, J. (2016). A multivariable prediction model for the chronification of non-traumatic shoulder pain: A systematic review. *Pain Physician, 19*(2), 1-10. https://doi.org/10.36076/ppj/2016.19.1

Sullivan, K. M., Dean, A., & Soe, M. M. (2009). OpenEpi: A web-based epidemiologic and statistical calculator for public health. *Public Health Reports 124*(3), 471-474. https://doi.org/10.1177/003335490912400320

Terwee, C. B., Bot, S. D., de Boer, M. R., van der Windt, D. A., Knol, D. L., Dekker, J., Bouter, L. M., & de Vet, H. C. (2007). Quality criteria were proposed for measurement

properties of health status questionnaires. *Journal of Clinical Epidemiology, 60*(1), 34-42. https://doi.org/10.1016/j.jclinepi.2006.03.012

Toemen, A., Dalton, S., & Sandford, F. (2011). The intra-and inter-rater reliability of manual muscle testing and a hand-held dynamometer for measuring wrist strength in symptomatic and asymptomatic subjects. *Hand Therapy, 16*(3), 67-74. https://doi.org/10.1258/ht.2011.01101

Tozzo, M. C., Ansanello, W., Martins, J., Zatiti, S. C. A., & de Oliveira, A. S. (2021). Inclinometer reliability for shoulder ranges of motion in individuals with subacromial impingement syndrome. *Journal of Manipulative and Physiological Therapeutics, 44*(3), 236-243. https://doi.org/10.1016/j.jmpt.2020.12.001

Treede, R.-D., Rief, W., Barke, A., Aziz, Q., Bennett, M. I., Benoliel, R., Cohen, M., Evers, S., Finnerup, N. B., First, M. B., Giamberardino, M. A., Kaasa, S., Kosek, E., Lavand'homme, P., Nicholas, M., Perrot, S., Scholz, J., Schug, S., Smith, B. H., Svensson, P., Vlaeyen, J. W. S., & Wang, S.-J. (2015). A classification of chronic pain for ICD-11. *Pain, 156*(6), 1003-1007. https://doi.org/10.1097/j.pain.0000000000000160

van Kampen, D. A., van den Berg, T., van der Woude, H. J., Castelein, R. M., Scholtes, V. A., Terwee, C. B., & Willems, W. J. (2014). The diagnostic value of the combination of patient characteristics, history, and clinical shoulder tests for the diagnosis of rotator cuff tear. *Journal of Orthopaedic Surgery and Research, 9*, 70. https://doi.org/10.1186/s13018-014-0070-y

van Riet, R. P., & Bell, S. N. (2011). Clinical evaluation of acromioclavicular joint pathology: Sensitivity of a new test. *Journal of Shoulder and Elbow Surgery, 20*(1), 73-76. https://doi.org/10.1016/j.jse.2010.05.023

Varacallo, M., El Bitar, Y., & Mair, S. D. (2023). Comprehensive Shoulder Evaluation Strategies. In *StatPearls*. StatPearls Publishing.

Vind, M., Bogh, S. B., Larsen, C. M., Knudsen, H. K., Søgaard, K., & Juul-Kristensen, B. (2011). Inter-examiner reproducibility of clinical tests and criteria used to identify subacromial impingement syndrome. *British Medical Journal Open, 1*(1), e000042. https://doi.org/10.1136/bmjopen-2010-000042

Walton, J., Mahajan, S., Paxinos, A., Marshall, J., Bryant, C., Shnier, R., Quinn, R., & Murrell, G. A. (2004). Diagnostic values of tests for acromioclavicular joint pain. *Journal of Bone Joint and Surgery, 86*(4), 807-812. https://doi.org/10.2106/00004623-200404000-00021

Walton, J., & Russell, S. (2015). Physiotherapy assessment of shoulder stiffness and how it influences management. *Shoulder & Elbow, 7*(3), 205-213. https://doi.org/10.1177/1758573215586152

White, S. G., McNair, P., Laslett, M., & Hing, W. (2015). Do patients undergoing physical testing report pain intensity reliably? *Arthritis Care and Research, 67*(6), 873-879. https://doi.org/10.1002/acr.22530

Willimon, S. C., Gaskill, T. R., & Millett, P. J. (2011, Feb). Acromioclavicular joint injuries: anatomy, diagnosis, and treatment. *Physician and Sportsmedicine, 39*(1), 116-122. https://doi.org/10.3810/psm.2011.02.1869

Wilson, E. B. (1927). Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association, 22*(158), 209-212. https://doi.org/10.2307/2276774

Woodland, L., Blignault, I., O'Callaghan, C., & Harris-Roxas, B. (2021). A framework for preferred practices in conducting culturally competent health research in a multicultural society. *Health Research Policy and Systems, 19*(1), 24. https://doi.org/10.1186/s12961-020-00657-y

Worksafe. (2019). *New Zealand Health and Safety at Work Strategy – Outcomes dashboard*. Retrieved 29 August 2021 from https://data.worksafe.govt.nz/editorial/work_related_health

World Health Organisation. (2021). *Musculoskeletal Conditions*. Retrieved 22 August 2021 from https://www.who.int/news-room/fact-sheets/detail/musculoskeletal-conditions

Yam, M. F., Loh, Y. C., Tan, C. S., Khadijah Adam, S., Abdul Manan, N., & Basir, R. (2018). General pathways of pain sensation and the major neurotransmitters involved in pain regulation. *International Journal of Molecular Sciences, 19*(8). https://doi.org/10.3390/ijms19082164

Yazigi, J. A., Nicolao, F. A., Matsunaga, F. T., Netto, N. A., Belloti, J. C., & Tamaoki, M. J. S. (2021). Supraspinatus tears: Predictability of magnetic resonance imaging findings based on clinical examination. *Journal of Shoulder and Elbow Surgery, 30*(8), 1834-1843. https://doi.org/10.1016/j.jse.2021.02.013

Appendices

Appendix A. Health and Disability Ethics Committees approval

Health and Disability Ethics Committees

Health and Disability Ethics Committees
Ministry of Health
133 Molesworth Street
PO Box 5013
Wellington
6011

hdecs@health.govt.nz

03 December 2020

Ms Monique Baigent
55 Parkhill Road Howick
Auckland 2014

Dear Ms Baigent,

Re:	**Ethics ref:**	**20/STH/203**
	Study title:	What is the reliability and diagnostic accuracy of individual and combinations of commonly used clinical tests for the diagnosis of chronic acromioclavicular joint pain?

I am pleased to advise that this application has been *approved* by the Southern Health and Disability Ethics Committee. This decision was made through the HDEC-Expedited Review pathway.

Conditions of HDEC approval

HDEC approval for this study is subject to the following conditions being met prior to the commencement of the study in New Zealand. It is your responsibility, and that of the study's sponsor, to ensure that these conditions are met. No further review by the Southern Health and Disability Ethics Committee is required.

Standard conditions:

1. Before the study commences at *any* locality in New Zealand, all relevant regulatory approvals must be obtained.
2. Before the study commences at *any* locality in New Zealand, it must be registered in a clinical trials registry. This should be a WHO-approved registry (such as the Australia New Zealand Clinical Trials Registry, www.anzctr.org.au) or https://clinicaltrials.gov/.
3. Before the study commences at *each given* locality in New Zealand, it must be authorised by that locality in Online Forms. Locality authorisation confirms that the locality is suitable for the safe and effective conduct of the study, and that local research governance issues have been addressed.

After HDEC review

Please refer to the *Standard Operating Procedures for Health and Disability Ethics Committees* (available on www.ethics.health.govt.nz) for HDEC requirements relating to amendments and other post-approval processes.

Your next progress report is due by 3 December 2021.

Participant access to ACC

The Southern Health and Disability Ethics Committee is satisfied that your study is not a clinical trial that is to be conducted principally for the benefit of the manufacturer or distributor of the medicine or item being trialled. Participants injured as a result of treatment received as part of your study may therefore be eligible for publicly-funded compensation through the Accident Compensation Corporation (ACC).

Please don't hesitate to contact the HDEC secretariat for further information. We wish you all the best for your study.

Yours sincerely,

Mrs. Helen Walker
Acting Chairperson
Southern Health and Disability Ethics Committee

Encl: appendix A: documents submitted
appendix B: statement of compliance and list of members

Appendix A
Documents submitted

Document	*Version*	*Date*
CV for CI	1	01 November 2020
Evidence of scientific review: Independent lecturer review by AUT the before proposal submission (Richard Ellis)	2	01 November 2020
Survey/questionnaire: Oxford shoulder score questionnaire	1	01 November 2020
Survey/questionnaire: SPADI questionnaire	1	01 November 2020
Survey/questionnaire: CSI questionnaire	1	01 November 2020
Survey/questionnaire: Standardised baseline assessment	1	01 November 2020
Survey/questionnaire: Patient waiting room handout	1	01 November 2020
Discussed additions for equity of Maori participants collated with the CMDHB Maori health team	5	01 November 2020
Mātauranga Māori Committee proposal AUT	3	01 November 2020
CMDHB Maori support letter	1	01 November 2020
PIS/CF: Participant consent form	3	01 November 2020
PIS/CF: Reliability study participant information sheet	4	01 November 2020
PIS/CF: Diagnostic accuracy study participant information sheet	4	01 November 2020
Protocol	24	04 November 2020
Application		04 November 2020
Declined letter for previous application in respect of the same (or substantially similar) study: Current version 28/11	5	28 November 2020
Declined letter for previous application in respect of the same (or substantially similar) study: current version 28/11/2020	5	28 November 2020
Covering Letter: Cover letter in response to feedback	2	28 November 2020
PIS/CF: amended consent form	4	28 November 2020
PIS/CF: amended version reliability study information sheet	5	28 November 2020
PIS/CF: amended version diagnostic accuracy study information sheet	5	28 November 2020
Protocol: Data management plan	1	28 November 2020
Response to Request for Further Information		

Statement of compliance and list of members

Statement of compliance

The Southern Health and Disability Ethics Committee:

- is constituted in accordance with its Terms of Reference
- operates in accordance with the *Standard Operating Procedures for Health and Disability Ethics Committees*, and with the principles of international good clinical practice (GCP)
- is approved by the Health Research Council of New Zealand's Ethics Committee for the purposes of section 25(1)(c) of the Health Research Council Act 1990
- is registered (number 00008713) with the US Department of Health and Human Services' Office for Human Research Protection (OHRP).

List of members

Name	*Category*	*Appointed*	*Term Expires*
Mrs Helen Walker	Lay (consumer/community perspectives)	19/08/2020	19/08/2021
Dr Pauline Boyles	Lay (consumer/community perspectives)	05/07/2019	05/07/2022
Dr Paul Chin	Non-lay (intervention studies)	27/10/2018	27/10/2021
Mr Dominic Fitchett	Lay (the law)	05/07/2019	05/07/2022
Dr Sarah Gunningham	Lay (other)	05/07/2016	05/07/2022
Assc Prof Mira Harrison-Woolrych	Non-lay (intervention studies)	28/06/2019	28/06/2020
Professor Jean Hay-Smith	Non-lay (health/disability service provision)	31/10/2018	31/10/2021
Dr Devonie Waaka	Non-lay (intervention studies)	18/07/2016	18/07/2019

Unless members resign, vacate or are removed from their office, every member of HDEC shall continue in office until their successor comes into office (HDEC Terms of Reference)

http://www.ethics.health.govt.nz

Appendix B. Counties Manukau Health research office approval

Research & Evaluation Office
Level 1, Ko Awatea, Middlemore Hospital
100 Hospital Road, Otahuhu; Private Bag 93311, Auckland – 1640
cmdhb.org.nz – koawatea.co.nz

03 February 2021

For the attention of: Monique Baigent

Thank you for the information you have supplied to the CM Health Research Office regarding the following research proposal:

CM Health Research Registration Number: 1330

Ethics Approval Reference Number: HDEC: 20/STH/203

Research Project Title: What is the reliability and diagnostic accuracy of individual and combinations of commonly used clinical tests for the diagnosis of chronic acromioclavicular joint pain?

I am pleased to inform you that the CM Health Research Office has received all the required service lead approvals and the Chief Medical Officer's final sign-off for the above research project, which has you named as the Principal Investigator.

This CM Health locality approval is valid until 5 February 2023, which is the Final Reporting Date specified on your registration information.

All external reporting requirements must be adhered to. Please note that failure to notify us of amendments, and/or submit copies of annual Progress Reports and annual Ethics renewal letters may result in the withdrawal of ethical and CM Health organisational approval.

FINAL REPORT: It is a requirement of the CM Health Research Policy that all research and audit projects conducted within CM Health should complete the CM Health Final Report Template and submit no later than 3 months following completion of the study. This report is to be uploaded to your study file on the Registry and is viewable by CM Health staff. Contact us for the report template or download it from the Registry. Please Note that having an overdue Final Report will impact your application for locality approval of any new studies.

Yours sincerely

Angela Bennett
Research Coordinator
Counties Manukau Health
Under delegated authority from CM Health Research Committee and the Chief Medical Officer

Appendix C. Mātauranga Māori Committee approval

School of Clinical Sciences
Mātauranga Māori Committee

Verification of Māori Consultation

This document provides verification that the research project named below was discussed with the School of Clinical Sciences Mātauranga Māori Committee, Auckland University of Technology. Specific comments and recommendations are indicated following the table.

Title of project: What is the diagnostic accuracy of individual and combinations of commonly used clinical tests for the diagnosis of chronic acromioclavicular joint pain?		
Research Team members and affiliations: Auckland University of Technology (AUT) and Counties Manukau District Health Board (CMDHB) Monique Baigent (AUT) Supervisors: Dr Steve White and Dr Julia Hill (AUT) Assessors: Michael Zo (physiotherapist), Mr Hogan Yeung (orthopaedic surgeon), Mr Brendan Coleman (orthopaedic surgeon) (CMDHB) Statistician: Assoc. Prof. Alain Vandal (CMDHB)		**Meeting Date:** 07/10/2020
Discussion Areas		**Discussed**
Whakapapa: Relationships		
Researcher experience in field		x
Consultation with local stakeholders		x
Consenting process		x
Clarity of data usage		x
Dissemination of findings		x
Benefits to participants		x
Protecting the rights & interest of Māori		
Clear purpose of project		X
Relevance to Māori		x
Likely outcome for participants, communities, other stakeholders		x
Participant recruitment methods		x
Māori involvement in project (participants, researchers, etc.)		x
Cultural & Social Responsibility		
Participants' access to appropriate advice		x
Participants treated with dignity and respect		x
Privacy and confidentiality		
Whānau support		x
Transparency of research process		x
Mana tangata – Power & Authority		

Reciprocity (acknowledgements, compensation, gifts)		x
Risks of participation identified		x
Ownership of outcomes		
Informed consent process		x

Comments

1.	Monique works at CMDHB across the outpatient and orthopedic knee/shoulder services and has been drawing on advice from Ulima Tofi from the Māori health team at CMDHB. This is a complex project being run within the orthopedic department aiming to address the issue of under-recognition and misdiagnosis of chronic acromioclavicular joint pain by trying to get a good diagnostic test / series of tests. Along-side this, the researchers are attempting to address equity of access issues raised by a recent CMDHB booking and scheduling review by adopting the Hui framework for all clinical communication and using the lead researcher to contact patients directly to make appointments rather than using the standard booking system.
2.	The movement of patients through the diagnostic accuracy testing pathway to diagnosis (and hence treatment) relies on the participants report of their level of pain during the testing process and any change reported after anesthetic injection into subacromial space.
3.	Cultural advisors at CMDHB will help facilitate the korero around experiences of Māori participants, and advise on the survey for those unable to attend to see how the processes used within the project may have impacted on issues identified in the earlier CHDHB work on engagement, and what can be carried forward into future service provision.
4.	In addition to the Hui process being drawn on, patients will be advised that whānau are welcome to attend their appointment and there will be an opportunity for whanau to have time and space for Karakia should they wish to. During the clinical appointment, permission will always be sought before progressing and focus will be on creating a context where patients feel comfortable.
4.	Summary of findings will be in Māori and English. Planning to outsource translations to official government service.

Recommendations made by Committee

1.	The committee applaud the thought and consideration that has been given to ensuring a culturally responsive approach. We do however recognise the complexity of what you are trying to achieve and acknowledge that this could easily be two separate projects. Given this, we promote caution and recommend you remain mindful that this project includes two separate but related pieces of work. It is important to ensure both components are given adequate thought and do not inadvertently become diluted. One way to ensure this is to more explicitly include research questions related to Māori equity and engagement in your research protocol. That said, we do understand that the primary research question for your qualification is the diagnostic accuracy component. Māori equity and engagement will be woven into the research design but we are unable to make it a research question without compromising the design. That being said our results and discussion will include our findings of Māori specific data and compare these results to non-Māori.

2.	Given importance of the pain rating scale to the success of this diagnostic testing, important that the pain rating scale is a culturally responsive measure for Māori. I have looked into this, however unfortunately no research has produced a Māori specific/tailored pain measure. It is worth having discussion around this in the korero "There are no culturally valid pain questionnaires for the Māori population." Hoeta, T., Baxter, G. D., Byrant, K., & Mani, R. (2020). Māori pain experiences and culturally valid pain assessment tools for Maori: A systematic narrative review. *New Zealand Journal of Physiotherapy*, 48(1): 37-50. Doi: 10.15619/NZJP/48.1.05
3.	It is good to see you offering time and space for Karakia. The committee suggests that you would not necessarily need to leave the room for that purpose, but rather than that whānau may welcome your collective involvement in Karakia. Further, while it would be good to invite Karakia you may find that whanau would welcome you leading a Karakia so it would be good to be prepared for that possibility. Many thanks for your suggestion. Our Māori health team have suggested a broad karakia in the way of Tūtawa which I will learn. Tūtawa mai i runga Tūtawa mai i raro Tūtawa mai i roto Tūtawa mai i waho Kia tau ai Te mauri tū Te mauri ora Ki te katoa Haumi e, Hui e, Tāiki e!
4.	We note that whānau will be expressly told that they are welcome to attend appointments. Ensure you give adequate time before the appointment for whānau to arrange for this should they wish to be present. Participants will be given 2 weeks notice minimum prior to their appointment.
5.	Be mindful of the need for simple language in information and summaries of findings. For the translation of material, there may be nuances around kupu used to convey health/scientific content. Further, there are some generational differences in Te Reo which may impact understanding of translations. It is worth getting translations checked by someone with a health-related background to ensure that participants are getting the understanding you are wanting them to take away. I will create the summary document in collaboration with the CMDHB Māori Health Team. If we do translate the full page of the research findings participants will be provided with both the Māori and English versions. I think your suggestion of perhaps not translating the whole document would be appropriate. If we translated only the introduction and our gratitude for their time, this may be easier to comprehend for a wider number of the participants.
7.	The current sampling process involves consecutive patients attending clinic. However, the committee queried how this would ensure Māori engagement in the research. The committee suggests you consider sampling consecutively for x number of Māori and consecutively for x number of non-Māori to ensure balance. For the diagnostic accuracy study, we suggest you oversample for Māori to ensure meaningful interpretation of findings for Māori. It is important to be mindful of the difference between having a representative sample (similar % as general

	population) and equal explanatory power (i.e. raising the statistical likelihood that findings are as relevant to Māori as non-Māori). We recognise this is challenging in achieve in practice, but recommend the intent is explicit in the protocol so that can inform targeted recruitment strategies if need be. The CMDHB Māori Research Reviewer was satisfied with our plan to uphold equity concerns within the study without oversampling Māori, and that such oversampling makes the study unfeasible in the setting of an MPhil research project. To ensure equivalent power in a Māori subgroup as in the original study, you would need to either only recruit Māori or double your sample size. In the first instance you *quintuple* the length of your recruitment period, in the second you multiply it by nearly seven.
8.	Whānau may not necessarily be aware of what the Hui process is that you are drawing on, or even be aware that you are trialing a new way of engaging with people before and during their appointments. We recommend you keep this in mind as you are refining protocols for the korero and survey. Thank you I will. We intend to ask more broad questions about whether they felt comfortable and listened to rather than talk about the Hui process.
9.	You may find it useful to critically look at your information sheet to make sure that there is a clear description of what participants can expect. You can also emphasise the invitation to bring support persons. We also recommend including photographs of the team in the information sheet so potential participants can put a face to the name, particularly given you will be engaging with them over the phone in advance of their appointment. I have added your suggestions and modified the information sheet thank you.

Please contact the Committee's Administrator Greta Smith at physioadmin@aut.ac.nz if you have any questions about this feedback, or any feedback of your own to let us know how things go.

You may be contacted in 12 months' time for feedback about the process and the usefulness of these comments and recommendations to your project.

Signature:  21/10/2020

Grant Mawston
Mātauranga Māori Consultation Committee

Appendix D. Counties Manukau Māori Health consultation

COUNTIES MANUKAU HEALTH

06/08/2020

Monique Baigent
Orthopaedic Physiotherapist
Middlemore Hospital
Counties Manukau DHB

RE: Maaori Consultation Request- Can clinical tests accurately diagnose chronic acromioclavicular joint (CACJ) pain?

Teenaa koe Monique

Thank you for asking me to review your study proposal. We met in person on 27th February which provided an opportunity for discussion. This meeting was followed up by several emails to clarify certain aspects of your study proposal. I want to acknowledge the effort you have invested into this process and in developing your personal knowledge around research with Maaori. I encourage you to utilise this knowledge in your future projects and practice.

This letter provides key points summary for the basis of decision

- You have addressed concerns regarding the production of appropriate levels of information for the different whaanau involved, the access of results, and the appropriate forums to disseminate study findings to ensure Maaori participants and communities do not miss out on valuable information.
- You have aligned your research with the Booking & Scheduling Review findings so to not perpetuate stand-alone research with no alignment to existing inequities.
- For Maaori who do suffer this injury this study has the potential to improve their outcome.

Based on the above points I endorse this study.

Ngaa Mihi,

Ulima Tofi
Ngaati Mamapoto, Rongowhakaata
Allied Health Lead- Maaori Health Development
Counties Manukau District Health Board

Appendix E. Australian New Zealand Clinical Trials Registry acceptance

Dear Monique Baigent,

Re: Can clinical tests accurately diagnose chronic acromioclavicular joint pain?

Thank you for submitting the above trial for inclusion in the Australian New Zealand Clinical Trials Registry (ANZCTR).

Your trial has now been successfully registered and allocated the ACTRN: ACTRN12621000348853

Web address of your trial:
https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?ACTRN=12621000348853
Date submitted: 27/01/2021 10:23:50 AM
Date registered: 26/03/2021 12:33:10 PM
Registered by: Monique Baigent
Principal Investigator: Monique Baigent

If you have already obtained Ethics approval for your trial, please send a copy of at least one Ethics Committee approval letter to mailto:info@actr.org.au or by fax to (+61 2) 9565 1863, attention to ANZCTR.

Note that updates should be made to the registration record as soon as any trial information changes or new information becomes available. Updates can be made at any time and the quality and accuracy of the information provided is the responsibility of the trial's primary sponsor or their representative (the registrant). For instructions on how to update please see https://www.anzctr.org.au/Support/HowToUpdate.aspx

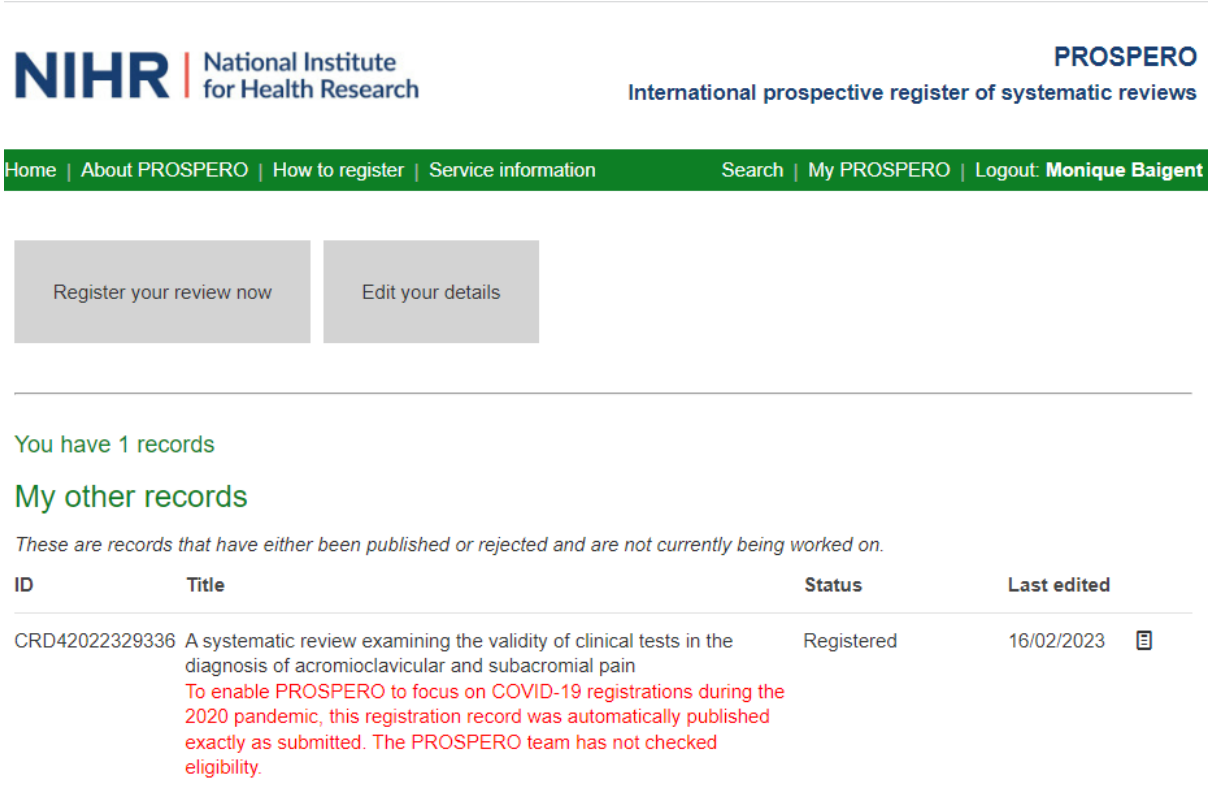
Please also note that the original data lodged at the time of trial registration and the tracked history of any changes made as updates will remain publicly available on the ANZCTR website.

The ANZCTR is recognised as an ICMJE acceptable registry (https://www.icmje.org/about-icmje/faqs/clinical-trials-registration/) and a Primary Registry in the WHO registry network (https://www.who.int/clinical-trials-registry-platform).

If you have any enquiries please send a message to info@actr.org.au or telephone +61 2 9562 5333.

Kind regards,
ANZCTR Staff
T: +61 2 9562 5333
F: +61 2 9565 1863
E: info@actr.org.au
W: https://www.anzctr.org.au/

Appendix F. PROSPERO registry record



NIHR | National Institute for Health Research

PROSPERO
International prospective register of systematic reviews

Home | About PROSPERO | How to register | Service information Search | My PROSPERO | Logout: **Monique Baigent**

Register your review now Edit your details

You have 1 records

My other records

These are records that have either been published or rejected and are not currently being worked on.

ID	Title	Status	Last edited
CRD42022329336	A systematic review examining the validity of clinical tests in the diagnosis of acromioclavicular and subacromial pain To enable PROSPERO to focus on COVID-19 registrations during the 2020 pandemic, this registration record was automatically published exactly as submitted. The PROSPERO team has not checked eligibility.	Registered	16/02/2023

Appendix G. Phone call dialogue for the reliability study

Call One

Hello my name is Monique Baigent I am calling from the orthopaedic service at the superclinic. We have received a referral from your GP about your shoulder pain. Is this still a problem for you? We are able to offer you an appointment with an orthopaedic clinician in a few weeks time. Would you be able to make.....

We will send you an appointment letter with your time and explaining where to come, over email and in the post (confirm addresses)

We are also running a research study to look at how we can improve our care for patients with shoulder problems. We will send you some information about it. The researcher involved will call you in 10 days time.

Call two:

Hello..... My name is Monique Baigent I am a researcher and physiotherapist calling from the Manukau Super Clinic. I just wanted to talk to you about your shoulder appointment and a research project we are running. Is this a good time to talk?

Have you received your appointment letter and the information about the research project? I am not sure if you have had time to read it but I would like to give you a general overview.

The goal of the research is to assess how helpful our shoulder tests are for you and other people. Two clinicians will assess your shoulder on the day and compare their results. It doesn't change how we manage your shoulder and the extra tests should help us treat you more effectively. I will meet you at your appointment and we will go over everything you need to know and answer any questions. We will also need extra time to do the testing approximately 90 minutes.

Would you be interested in being part of this study? Do you have any questions?

NO: Thats perfectly fine we will see you for your appointment time you won't miss out in any way. Thank you for your time see you on the.... at .....

Yes: To save some time for you on the day could I ask you a few questions about your health?

- Do you have a current or previous history of cancer?
- Do u have rheumatoid arthritis, (swelling of the fingers and toes)
- Is it your left or right shoulder?
- Do you have a current or previous shoulder infection, fractures or dislocations on your .....side
- Have you had previous shoulder or neck surgery on your ..... side
- have you had any previous shoulder injections on your ..... side

On your file it says that you are taking.....Is this still the case? How often are you taking.....?

If intermittent: could we ask you to not take your medication on the day of your appointment as it may confuse your testing

If regular: Please could you take your medication as you normally would on the day of your appointment

We will see you at your appointment, you are welcome to bring your whanau or support person if you would like to. If you decide not to be in the study we will still see you at your appointment time as normal and you will not be disadvantaged in any way.

Appendix H. Phone call dialogue for the diagnostic accuracy study

Call One

Hello my name is ..... I am calling from the orthopaedic service at the superclinic. We have received a referral from your GP about your shoulder pain. Is this still a problem for you? We are able to offer you an appointment in a few weeks time. Would you be able to make.....

We will send you an appointment letter explaining where to come over email and in the post? (confirm addresses)

We are also running a research study to look at how we can improve our care for patients with shoulder problems. We will send you some information about it. The researcher involved will call you in 10 days time.

Call two:

Hello..... My name is ..... I am a researcher and physiotherapist calling from the Manukau Super Clinic

We called you just over a week ago and sent you some information about a research project

Did you get a chance to read any of that information?

YES: Did you have any questions about the research?

Would you be interested in being part of this study?

NO: Thats ok, its about a project we are research running to help us improve our tests for the shoulder. Would you be interested in finding out more?

YES: It would be a similar process to your normal shoulder assessment, however we would record your testing and instead of a 30minute appointment it would be 1 hour and 30 minutes. This would include time to do more testing and to inject the area the pain is likely to be coming from. There is also a chance you may be asked to attend a second appointment at the radiology department for a different injection to tell us more information about your shoulder. Do you have any questions?

Would you be interested in being part of this study?

NO: Thats perfectly fine we will see you for your appointment time you won't miss out in any way. Thank you for your time see you on the.... at .....

To save some time for you on the day could I ask you a few questions about your health?

- Do you have a current or previous history of cancer?
- Do u have rheumatoid arthritis, (swelling of the fingers and toes)
- Is it your left or right shoulder?
- Do you have a current or previous shoulder infection, fractures or dislocations on your .....side
- Have you had previous shoulder or neck surgery on your ..... side
- have you had any previous shoulder injections on your ..... side

Could we also ask for you to not take any pain medication for 6 hours before your appointment?

We will see you at your appointment, you are welcome to bring your whanau or support person if you would like to. If you decide not to be in the study we will still see you at your appointment time as normal and you won't miss out in any way.

Appendix I. Participant consent form

Consent Form

Project title: **How can we improve our testing for shoulder joint pain?**

Project Supervisor: ***Dr Steve White***

Researcher: ***Monique Baigent***

- ○ I have read and understood the information provided about this research project in the Information Sheet dated 15/09/2021.
- ○ I have been given an opportunity to ask questions and to have them answered.
- ○ I understand that notes will be taken during the assessment and this data will be stored on the lead researcher's password protected computer.
- ○ I understand that taking part in this study is my choice and that I may leave the study at any time without being disadvantaged in any way.
- ○ I understand that if I leave the study then I will be offered the choice of having data removed or allowing it to continue to be used. However, once the findings have been produced, removal of my data may not be possible.
- ○ I agree to take part in this research.
- ○ I consent for my de-identified data from this study to be used in future studies
- ○ I wish to receive a summary of the research findings (please tick one): Yes○ No○
- ○ For Māori patients: I would like to be part of a group korero following the study (please tick one): Yes○ No○

Participant signature: .....

Participant name: .....

Participant contact details :

.....

.....

.....

.....

Date:

Approved by the Health and Disability Ethics Committee on 03/12/2020, HDEC Reference number 20/STH/203

Note: The Participant should retain a copy of this form.

Appendix J. Reliability study participant information sheet

Participant Information Sheet- Physiotherapist

Date Information Sheet Produced:

13/05/2020

Project Title

How can we improve our testing for shoulder joint pain?

An Invitation

Kia Ora/Hello

My name is Monique Baigent and I am undertaking research as a part of my Masters of Philosophy qualification at Auckland University of Technology. My research is focussed on the diagnosis of problems with a joint in the shoulder known as the acromioclavicular joint (ACJ). These problems are often missed as they are not easy to diagnose in the clinic and can be confused with another common cause of shoulder pain known as subacromial pain. I would like to invite you to participate in this research

What is the purpose of this research?

We are testing the reliability of clinical tests to identify a common shoulder pathology known as acromioclavicular joint (ACJ) pain. We are inviting patients who have been referred to Counties Manukau District Health Board (CMDHB) orthopaedic clinic with symptoms that suggest ACJ or sub-acromial pain to participate.

The clinical tests and x-rays and scans we currently use are not accurate enough to diagnose these disorders confidently. This study will improve our knowledge in this area and may lead to better management of ACJ pain. The findings of this research may be used for academic publications and presentations.

How was I identified and why am I being invited to participate in this research?

Your referral for shoulder pain to the orthopaedic clinic has information from your doctor that indicates you may have ACJ or sub-acromial pain. Participating in this study will allow us to be more clear of your diagnosis and offer you more specific treatment options.

How do I agree to participate in this research?

We may ask you to participate in the study after we have examined/assessed you in the clinic. If you meet the criteria for the study, we will then formally ask you if you would like to be part of it. You will have the opportunity to ask any questions about the study. If you agree to participate, you will be required to sign a consent form. This form will then be stored in the research supervisor's office at AUT in a secure manner.

Your participation in this research is entirely voluntary. You can decline to participate or withdraw from the study at any time prior to the completion of data collection. If you choose to withdraw from the study, then you will be offered the choice between having any data belonging to you removed or allowing it to continue to be used. However, once the findings have been produced, removal of your data may not be possible.

What will happen in this research and where will this take place?

- Participants will attend their appointment at Module one Manukau Super Clinic. We would like to extend the invite for your whanau or support persons should you wish to bring them.
- When you arrive check in at reception and the clinician will come to meet you, introduce themselves and talk more about the research.
- The clinician will ask you questions about your shoulder and then complete some physical tests to help with the diagnosis.
- If you meet the study criteria and you are happy to be in the study, the clinician will ask you for the signed consent form. If you choose not to be part of the study at any point you will still receive standard care of assessment and treatment.
- After 30 minutes your shoulder will be reassessed by another clinician and the results will be compared. You will then be offered management options for your shoulder condition.
- We are looking to measure how the testing and results vary between clinicians to establish the consistency of testing. This will help to guide what tests clinicians use and in what order.

What are the discomforts and risks and how will they be alleviated?

During the clinical assessment the physical testing may reproduce your shoulder symptoms. In the unlikely event that you are injured in this study, you would be eligible to apply for compensation from ACC just as you would be if you were injured in an accident at work or at home. This does not mean that your claim will automatically be accepted. You will have to lodge a claim with ACC, which may take some time to assess. If your claim is accepted, you will receive funding to assist in your recovery.

Everything that you say will be treated with the highest confidence. All possible steps will be taken to protect your confidentiality. You will be able to withdraw from the project at any stage. Participation is voluntary. You can refuse to participate at any time without any penalty.

What are the benefits?

The main potential benefit for participants in this study is to receive a correct diagnosis for their long-standing pain. You will receive a full thorough assessment from two clinicians as well as your treatment. This help to confirm your diagnosis and allow the most appropriate treatment plan to be developed.

The results of this research will be also be valuable to physiotherapists, doctors and surgeons assessing the shoulder joint and help future patients receive more targeted and appropriate care. Your input will also contribute towards the completion of my Masters of Philosophy thesis.

How will my privacy be protected?

During this study the surgeon and/or physiotherapist will record information about you and your study participation. This includes the results of any study assessments. If needed, information from your hospital records and your GP may also be collected. You cannot take part in this study if you do not consent to the collection of this information.

Identifiable Information

Identifiable information is any data that could identify you (e.g. your name, date of birth, or address). Only researchers and clinical staff at the Manukau Superclinic will have access to your identifiable information.

- Your surgeon and/or physiotherapist assessing you
- radiology department staff, to process and report your tests
- The ethics committees or government agencies from New Zealand if the study or site is audited. Audits are done to make sure that participants are protected, the study is run properly, and the data collected is correct.
- Your family doctor will receive a summary of your testing results

De-identified (Coded) Information

To make sure your personal information is kept confidential, information that identifies you will not be included in any report generated by the researcher. Instead, you will be identified by a code. The researcher will keep a list linking your code with your name, so that you can be identified by your coded data if needed.

The results of the study may be published or presented, but not in a form that would reasonably be expected to identify you.

Future Research Using Your Information.

If you agree, your coded information may be used for future research related to shoulder pain.

You will not be told when future research is undertaken using your information. Your information will only be used by the researchers in this study. You will not get reports or other information about future research that is done outside this study using your information.

Security and Storage of Your Information.

Your identifiable information is held at Middlemore Hospital during the study. After the study it is transferred to a secure archiving site and stored for at least 10 years, then destroyed. Your coded information will be entered into electronic case report forms and sent through a secure server to the supervisor of this project at AUT. Coded study information will be kept by the supervisor in secure, cloud-based storage indefinitely. All storage will comply with local and/or international data security guidelines.

Risks.

Although efforts will be made to protect your privacy, absolute confidentiality of your information cannot be guaranteed. Even with coded and anonymised information, there is no guarantee that you cannot be identified. The risk of people accessing and misusing your information (e.g. making it harder for you to get or keep a job or health insurance) is currently very small, but may increase in the future as people find new ways of tracing information.

Rights to Access Your Information.

You have the right to request access to your information held by the research team. You also have the right to request that any information you disagree with is corrected.

Please ask if you would like to access the results of your screening and safety tests during the study.

If you have any questions about the collection and use of information about you, you should ask the researcher or research supervisor (contacts below).

Rights to Withdraw Your Information.

You may withdraw your consent for the collection and use of your information at any time, by informing your study clinician.

If you withdraw your consent, your study participation will end, and the study team will stop collecting information from you.

If you agree, information collected up until your withdrawal from the study will continue to be used and included in the study. You may ask for it to be deleted when you withdraw, unless you withdraw after the study analyses have been undertaken.

What are the costs of participating in this research?

The cost to you of participating in this research will essentially be your time. The initial assessment will take approximately 90 minutes, compared to the standard appointment time which is 30 minutes.

What opportunity do I have to consider this invitation?

If you are interested in participating in this study please let your clinician know when you attend your appointment. Please remember that your participation is voluntary. If you need further information or clarification of any aspect of the project, please contact myself or my supervisor. Contact details are given below.

Will I receive feedback on the results of this research?

Yes, if you wish, you will receive a summary of the study. You may also receive a copy of any papers that are generated from this study on request.

What do I do if I have concerns about this research?

Any concerns regarding the nature of this project should be notified in the first instance to the Project Supervisor, Dr Steve White, phone: 09 921 9999 ext 7073 email: Steve.white@aut.ac.nz

We also have a Maori research support person you may contact with questions Te Hao Apaapa-Timu, phone: 021 576 276 email: TeHao.Apaapa-Timu@middlemore.co.nz

Whom do I contact for further information about this research?

Please keep this Information Sheet and a copy of the Consent Form for your future reference. You are also able to contact the research team as follows:

Researcher Contact Details:

Monique Baigent

Email: monique.baigent@middlemore.co.nz

Phone: 0210510780

Project Supervisor Contact Details:

Dr Steve White

Email: Steve.white@aut.ac.nz

Phone 09 921 9999 ext 7073

If you want to talk to someone who isn't involved with the study, you can contact an independent health and disability advocate on:

Phone:	0800 555 050
Fax:	0800 2 SUPPORT (0800 2787 7678)
Email:	advocacy@advocacy.org.nz
Website:	https://www.advocacy.org.nz/

You can also contact the health and disability ethics committee (HDEC) that approved this study on:

Phone:	0800 4 ETHIC
Email:	hdecs@health.govt.nz

Approved by the Health and Disability Ethics Committee on 03/12/2020, HDEC Reference number 20/STH/203

Appendix K. Diagnostic accuracy study participant information sheet

Participant Information Sheet

Date Information Sheet Produced:

15/09/2021

Project Title

How can we improve our testing for shoulder joint pain?

An Invitation

Kia Ora/Hello

My name is Monique Baigent and I am currently undertaking research as a part of my Masters of Philosophy qualification at Auckland University of Technology. My research is focussed on the diagnosis of problems with a joint in the shoulder known as the acromioclavicular joint (ACJ). These problems are often missed as they are not easy to diagnose in the clinic and can be confused with another common cause of shoulder pain known as subacromial pain. I would like to invite you to participate in this research

What is the purpose of this research?

We are testing the usefulness of clinical tests to identify a common shoulder pathology known as acromioclavicular joint (ACJ) pain. We are inviting patients who have been referred to Counties Manukau District Health Board (CMDHB) orthopaedic clinic with symptoms that suggest ACJ or sub-acromial pain to participate.

The clinical tests and x-rays and scans we currently use are not accurate enough to diagnose these disorders confidently. This study will improve our knowledge in this area and may lead to better management of ACJ pain. The findings of this research may be used for academic publications and presentations.

How was I identified and why am I being invited to participate in this research?

Your referral for shoulder pain to the orthopaedic clinic has information from your doctor that indicates you may have ACJ or sub-acromial pain. Participating in this study will allow us to be more clear of your diagnosis and offer you more specific treatment options.

How do I agree to participate in this research?

We may ask you to participate in the study after we have tested you in the clinic. If you meet the criteria for the study, we will formally ask you if you would like to be part of it. You will have the opportunity to ask any questions that you may have about the study. If you agree to participate, you will be required to sign a consent form. This form will then be stored in the research supervisor's office at AUT in a secure manner.

Your participation in this research is entirely voluntary. You can decline to participate or withdraw from the study at any time prior to the completion of data collection. If you choose to withdraw from the study, then you will be offered the choice between having any data belonging to you removed or allowing it to continue to be used. However, once the findings have been produced, removal of your data may not be possible.

What will happen in this research and where this will take place?

Participants will attend their appointment at Module one Manukau Super Clinic. We would like to extend the invite for your whanau or support persons should you wish to bring them. When you arrive please check in with the receptionist and the clinician will come to get you, introduce themselves and talk more about the research.

The clinician will ask you questions about your shoulder then go through some physical tests to help with the diagnosis. If you meet the criteria and you are happy to be in the study the clinician will ask you for the signed consent form. If you choose not to be part of the study at any point you will still receive the standard care of assessment and treatment.

Participants will then be given an anaesthetic injection into their sub-acromial space. This is the best test the medical profession has to diagnose pain coming from the rotator cuff. Thirty minutes following the injection three of the physical tests will be repeated. If you have a positive response to this injection then you will be offered treatment options for sub-acromial pain.

Should there be little or no relief, it is likely you have ACJ pain and the clinician will arrange for you to have an anaesthetic injection into this joint. When you arrive at this second appointment at Middlemore Hospital radiology a physiotherapist will be there to test your shoulder before and after your injection. Once your response to the injection is known, the clinician can advise and offer you the most appropriate treatment options.

Under normal practice in the clinic patients would be offered a combined steroid and anaesthetic injection. Your injection will only contain anaesthetic however if one of these injections relieves your pain you will be offered a steroid injection following if you wish.

What are the discomforts and risks and how will they be alleviated?

During the clinical assessment the physical testing may reproduce your shoulder symptoms. In the unlikely event that you are injured in this study, you would be eligible to apply for compensation from ACC just as you would be if you were injured in an accident at work or at home. This does not mean that your claim will automatically be accepted. You will have to lodge a claim with ACC, which may take some time to assess. If your claim is accepted, you will receive funding to assist in your recovery.

There are some minor known side effects of administering an aesthetic injection. These are as follows:

- Mild discomfort associated with insertion of the needle
- A tingling sensation as the medication wears off
- Minor bruising, bleeding or soreness where the injection was given.

In very rare cases with high anaesthetic doses there is a small chance of:

- Seizures
- Lowered blood pressure

- Slowed heart rate
- Breathing problems
- Allergic reactions
- Infection from the needle site

Your dose will be very low and the injection will be done under sterile conditions. This is a short lasting anaesthetic which should wear off in approximately 4 hours. In the time before the anaesthetic wears off participants will be advised to not lift anything heavy or perform activities that normally aggravate for up to four hours afterwards to prevent injury. The injection will be provided in medical facilities (Middlemore Hospital and Counties Manukau Super Clinic) where trained doctors and nurses are onsite to available should a participant need assistance.

What are the benefits?

The main potential benefit for participants in this study is to reach a correct diagnosis for their long-standing pain. An anaesthetic injection is the best tool we have available to diagnose shoulder pain. Participants in this study will have immediate access to an injection at no cost. This will confirm the participant's diagnosis and allow the most appropriate treatment plan to be developed.

The results of this research will be also be valuable to physiotherapists, doctors and surgeons assessing the shoulder joint and help future patients receive more targeted and appropriate care. Your input will also contribute towards the completion of my Masters of Philosophy thesis.

How will my privacy be protected?

During this study the surgeon and/or physiotherapist will record information about you and your study participation. This includes the results of any study assessments. If needed, information from your hospital records and your GP may also be collected. You cannot take part in this study if you do not consent to the collection of this information.

<u>Identifiable Information</u>

Identifiable information is any data that could identify you (e.g. your name, date of birth, or address). Only researchers and clinical staff at the Manukau Superclinic or Middlemore hospital will have access to your identifiable information.

- Your surgeon and/or physiotherapist assessing you
- radiology department staff, to process and report your tests
- The ethics committees or government agencies from New Zealand if the study or site is audited. Audits are done to make sure that participants are protected, the study is run properly, and the data collected is correct.
- Your family doctor will receive a summary of your testing results

<u>De-identified (Coded) Information</u>

To make sure your personal information is kept confidential, information that identifies you will not be included in any report generated by the researcher. Instead, you will be identified by a

code. The researcher will keep a list linking your code with your name, so that you can be identified by your coded data if needed.

The results of the study may be published or presented, but not in a form that would reasonably be expected to identify you.

Future Research Using Your Information.

If you agree, your coded information may be used for future research related to shoulder pain.

You will not be told when future research is undertaken using your information. Your information will only be used by the researchers in this study. You will not get reports or other information about future research that is done outside this study using your information.

Security and Storage of Your Information.

Your identifiable information is held at Middlemore Hospital during the study. After the study it is transferred to a secure archiving site and stored for at least 10 years, then destroyed. Your coded information will be entered into electronic case report forms and sent through a secure server to the supervisor of this project at AUT. Coded study information will be kept by the supervisor in secure, cloud-based storage indefinitely. All storage will comply with local and/or international data security guidelines.

Risks.

Although efforts will be made to protect your privacy, absolute confidentiality of your information cannot be guaranteed. Even with coded and anonymised information, there is no guarantee that you cannot be identified. The risk of people accessing and misusing your information (e.g. making it harder for you to get or keep a job or health insurance) is currently very small, but may increase in the future as people find new ways of tracing information.

Rights to Access Your Information.

You have the right to request access to your information held by the research team. You also have the right to request that any information you disagree with is corrected.

Please ask if you would like to access the results of your screening and safety tests during the study.

If you have any questions about the collection and use of information about you, you should ask the researcher or research supervisor (contacts below).

Rights to Withdraw Your Information.

You may withdraw your consent for the collection and use of your information at any time, by informing your study clinician.

If you withdraw your consent, your study participation will end, and the study team will stop collecting information from you.

If you agree, information collected up until your withdrawal from the study will continue to be used and included in the study. You may ask for it to be deleted when you withdraw, unless you withdraw after the study analyses have been undertaken.

What are the costs of participating in this research?

The cost to you of participating in this research will essentially be your time. The initial assessment will take approximately 90 minutes, compared to the standard appointment time which is 30 minutes. If you require an injection into your acromioclavicular joint you will be asked to attend for a second appointment at Middlemore Hospital radiology for approximately one hour. If you are required to travel to Middlemore a koha in the form of a $20 petrol voucher will be provided to cover petrol and parking costs.

What opportunity do I have to consider this invitation?

If you are interested in participating in this study please let me or your clinician know when you attend your appointment. Please remember that your participation is voluntary. If you need further information or clarification of any aspect of the project, please contact myself or my supervisor. Contact details are given below.

Will I receive feedback on the results of this research?

Yes, if you so wish, you will receive a summary of the study. You may also receive a copy of any papers that are generated from this study on request.

What do I do if I have concerns about this research?

Any concerns regarding the nature of this project should be notified in the first instance to the Project Supervisor, Dr Steve White, phone: 09 921 9999 ext 7073 email: Steve.white@aut.ac.nz

We also have a Maori research support person you may contact with questions Te Hao Apaapa-Timu, phone: 021 576 276 email: TeHao.Apaapa-Timu@middlemore.co.nz

Whom do I contact for further information about this research?

You are also able to contact the research team as follows:

Researcher Contact Details:

Monique Baigent

Email: monique.baigent@middlemore.co.nz

Phone: 0210510780

Project Supervisor Contact Details:

Dr Steve White

Email: Steve.white@aut.ac.nz

Phone 09 921 9999 ext 7073

Please keep this Information Sheet and a copy of the Consent Form for your future reference.

If you want to talk to someone who isn't involved with the study, you can contact an independent health and disability advocate on:

Phone:	0800 555 050
Fax:	0800 2 SUPPORT (0800 2787 7678)
Email:	advocacy@advocacy.org.nz
Website:	https://www.advocacy.org.nz/

You can also contact the health and disability ethics committee (HDEC) that approved this study on:

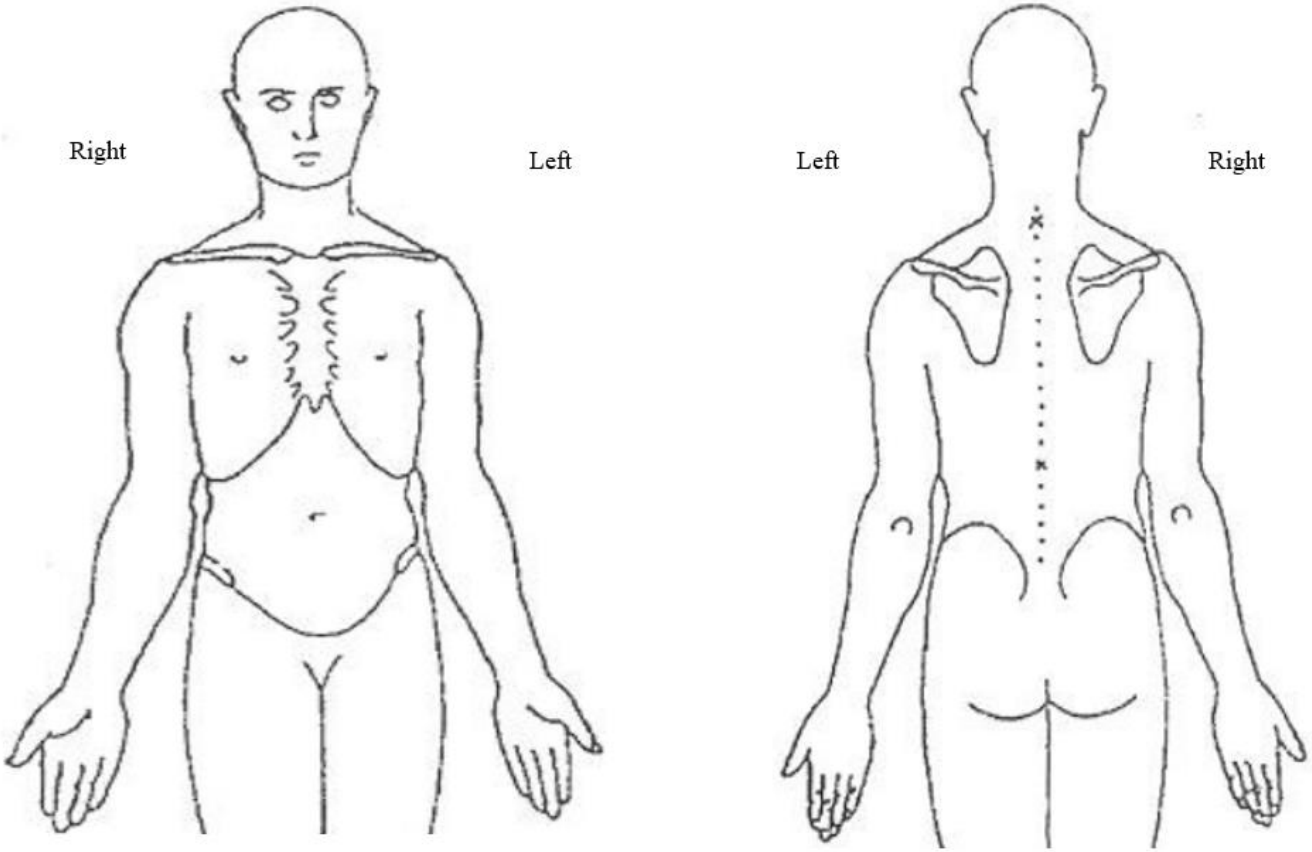
Phone:	0800 4 ETHIC
Email:	hdecs@health.govt.nz

Approved by the Health and Disability Ethics Committee on 03/12/2020, HDEC Reference number 20/STH/203

Appendix L. Participant pain diagram

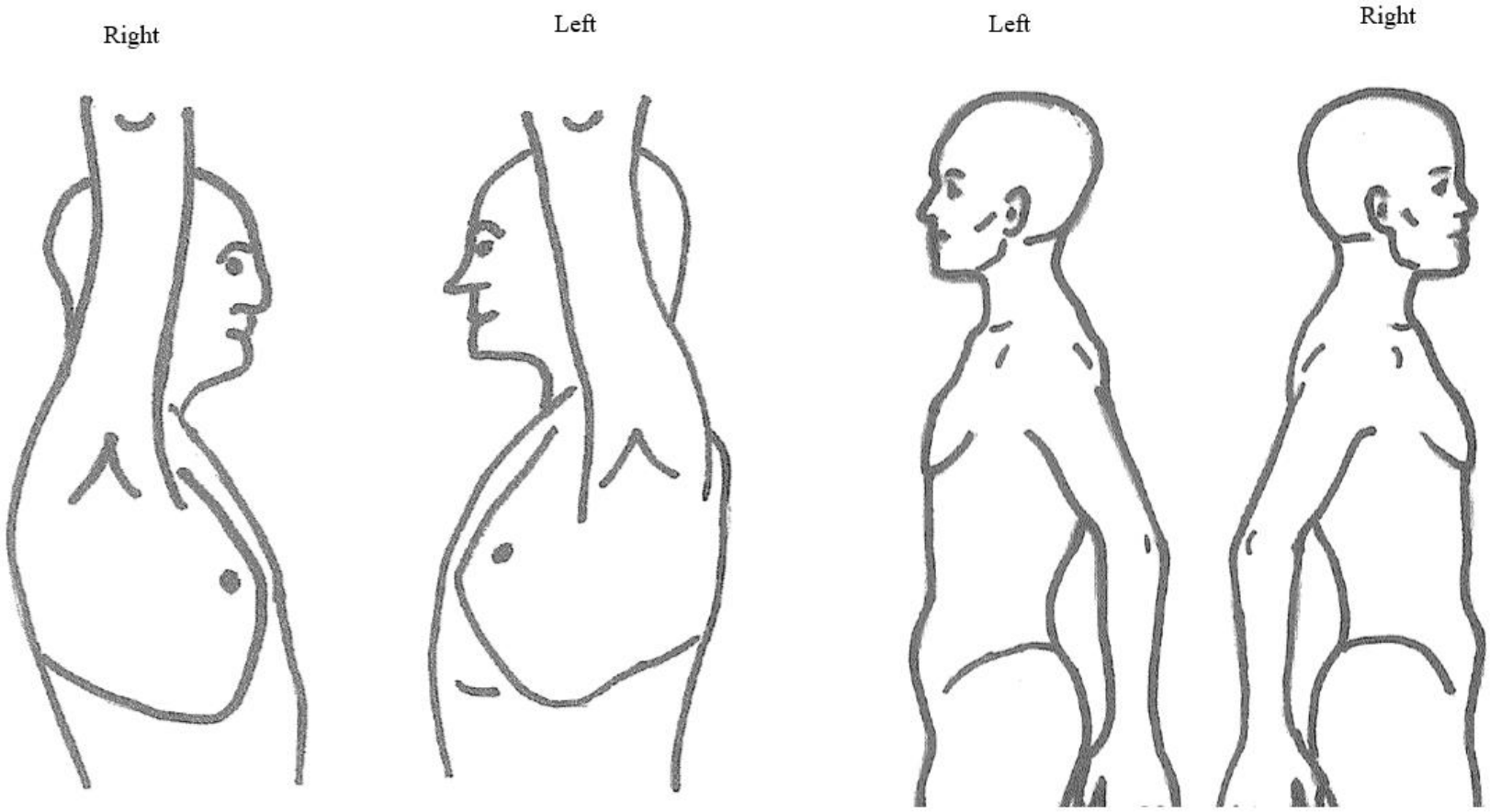
Please draw your pain location on the pain diagram.
Draw a small red X over spot of your worst pain.
Colour pain in green, any pins and needles in blue and any numbness in purple

COUNTIES MANUKAU HEALTH
Healthy Together



Please draw your pain location on the pain diagram.
Draw a small red X over spot of your worst pain.
Colour pain in green, any pins and needles in blue and any numbness in purple

COUNTIES MANUKAU HEALTH
Healthy Together



Appendix M. Central Sensitisation Inventory

Worksheet

CSI Inventory (Part A)

Name _____ Date _____

Please circle the best response to the right of each statement.
Key for Scoring: **Never = 0, Rarely = 1, Sometimes = 2, Often = 3, Always = 4**

1. I feel tired and unrefreshed when I wake from sleeping.	Never	Rarely	Sometimes	Often	Always
2. My muscles feel stiff and achy.	Never	Rarely	Sometimes	Often	Always
3. I have anxiety attacks.	Never	Rarely	Sometimes	Often	Always
4. I grind or clench my teeth.	Never	Rarely	Sometimes	Often	Always
5. I have problems with diarrhea and/or constipation.	Never	Rarely	Sometimes	Often	Always
6. I need help in performing my daily activities.	Never	Rarely	Sometimes	Often	Always
7. I am sensitive to bright lights.	Never	Rarely	Sometimes	Often	Always
8. I get tired very easily when I am physically active.	Never	Rarely	Sometimes	Often	Always
9. I feel pain all over my body.	Never	Rarely	Sometimes	Often	Always
10. I have headaches.	Never	Rarely	Sometimes	Often	Always
11. I feel discomfort in my bladder and/or burning when I urinate.	Never	Rarely	Sometimes	Often	Always
12. I do not sleep well.	Never	Rarely	Sometimes	Often	Always
13. I have difficulty concentrating.	Never	Rarely	Sometimes	Often	Always
14. I have skin problems such as dryness, itchiness, or rashes.	Never	Rarely	Sometimes	Often	Always
15. Stress makes my physical symptoms get worse.	Never	Rarely	Sometimes	Often	Always
16. I feel sad or depressed.	Never	Rarely	Sometimes	Often	Always
17. I have low energy.	Never	Rarely	Sometimes	Often	Always
18. I have muscle tension in my neck and shoulders.	Never	Rarely	Sometimes	Often	Always
19. I have pain in my jaw.	Never	Rarely	Sometimes	Often	Always
20. Certain smells, such as perfumes, make me feel dizzy and nauseated.	Never	Rarely	Sometimes	Often	Always
21. I have to urinate frequently.	Never	Rarely	Sometimes	Often	Always
22. My legs feel uncomfortable and restless when I am trying to go to sleep at night.	Never	Rarely	Sometimes	Often	Always
23. I have difficulty remembering things.	Never	Rarely	Sometimes	Often	Always
24. I suffered trauma as a child.	Never	Rarely	Sometimes	Often	Always
25. I have pain in my pelvic area.	Never	Rarely	Sometimes	Often	Always
Total Each Column					
				Overall Total	

Copyright © 2017, Martha Teater & Don Teater. *Treating Chronic Pain*. All rights reserved.

Appendix N. Oxford Shoulder Score

PROBLEMS WITH YOUR SHOULDER

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
☐	☐	☐	☐	☐

© Isis Innovation Limited, 1998. All rights reserved. Oxford Shoulder Score – English for the United Kingdom

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
☐	☐	☐	☐	☐

Finally, please check back that you have answered each question.
Thank you very much.

© Isis Innovation Limited, 1998. All rights reserved. Oxford Shoulder Score – English for the United Kingdom

Appendix O. Standardised interview assessment

Standardised Assessment Data Collection Form

	Positive	Negative	Mechanism
Subjective assessment			
What caused your pain?			
1. An injury? (external force, fall or impact)			
2. A strain (stretching, reaching or lifting)			
3. A job, sport or activity over time?			
4. No known cause?			
How long have you had this pain (months)?			
What is your pain score at worst in the last week (0 no pain 10 the worst imaginable)			
Is your pain there every minute of the day?			
If yes what is the lowest level of pain you experience? (out of 10)			
With one finger point to where it hurts the most? (dominant pain)			
Do you have pain below your elbow? (secondary pain)			
Do you have pain into your hand? (secondary pain)			
Do you have neck pain? (secondary pain)			
Do you have any pins and needles on your affected arm?	Where? Intermittent or constant?		
Do you have any numbness on your affected arm?	Where? Intermittent or constant?		
Are you left or right handed?			
Which shoulder is affected?			
Do you have trouble lying on your sore side?			
Do you have trouble lying on your other side due to your shoulder pain?			
Does your shoulder pain keep you up at night?			
If yes is your shoulder pain worse at night than during the day?			
	Positive	Negative	
Do you have difficulty with overhead tasks?			
Do you have difficulty bringing your hand up your back? demonstrate			
Do you have difficulty with moving your arm behind you? demonstrate			
Do you have difficulty with reaching across your chest? demonstrate			
Are there any clicking noises when you move your arm?			
Are you a smoker?			
Have you ever been a smoker?			
If so for how many years?			
Do you play any sports?			
If not have you had to give up because of your shoulder?			
Are you working?			
If yes: are you able to do all of your normal work tasks?			

Are you not working because of your shoulder pain?		
Have you had a cortisone injection into this shoulder?	How many?	
Do you know where in your shoulder they put it?		
When did you have the injection?		
Was it helpful for the first few hours afterwards?		
Did it relieve all of your pain? (%)		
How long did the relief last for?		
X-ray evidence of OA or osteolysis **Check x-ray to rule out Severe GHJ OA (3 or more on the Kellgren Lawrence scale) or calcific tendinitis.**		
MRI findings of bony oedema, joint cysts, capsular hypertrophy and/or osteophytes		

Appendix P. Reliability study standardised physical assessment

"Tell me when the pain starts to come on for these tests and I will stop"
"show me where the pain is and how severe it is out of 10"(while performing the test)
Tests can be done more than once to ensure they are accurate.
If a test is painfree apply overpressure.
Resisted testing: the participant pushes against the examiner until pain or max pressure.
Post injection test the 3 most painful tests first (if more than 3 highest pick the first 3)

Objective assessment	Positive (provocative of your familiar/dominant pain)	Negative (not their dominant pain or less < 2/10 pain) Record 2° pain	Pain response (out of 10) Maintain the test while you ask them	Unable to complete the test (why)	Indeterminate result (unclear response)
1.Patient generated provocative movement: Can you show me the most painful movement for your arm?	Describe the position:				
2. Tenderness over ACJ (test both sides at once)					
3. Cross arm adduction stress test (adduct first then IR)					
4.External rotation adduction test					
5. AC resisted extension (resisted horizontal abduction),					
6. Active compression test (O'Briens) positive if external rotation is less painful			Internal rotation: External rotation:		
7. Paxinos Sign (two hands as needed)					
8. Bell Van Reit test (start in ER then wind up)					
9. Hawkins Kennedy Test					
10. Pain with					

Scapula protraction (overpressure on humeral head)					
10. Pain with Scapula retraction (overpressure on humeral head)					
10. Pain with Scapula elevation (overpressure on humeral head)					
10. Pain with Scapula depression (overpressure on humeral head)					

Shoulder active range of motion (affected arm only)	Range of motion with an inclinometer	Reproduction of their primary pain (NPRS out of 10)	Unable to complete the test	Indeterminate result (unclear response)
11. Flexion				
11. Scaption				Painful arc Y/N
11. External rotation	N/A			
11. Internal rotation (the level of spinous process the participant can reach up their back)				

Resisted testing (compare to other side)	Strength out of 5 (Oxford Scale)	Reproduction of their primary pain (NPRS out of 10)	Unable to complete the test	Indeterminate result (unclear response)
12. Flexion 10degrees				
12. Abduction 10 degrees				
12. External rotation (hand at side in neutral and the clinicians hand between the elbow and their side)				
12. Belly Press (arm at side if unable)				

Passive range of motion	Range of motion with an inclinometer	Reproduction of their primary pain (NPRS out of 10)	Unable to complete the test	Indeterminate result (unclear response)
13. Flexion				
13. Abduction				
13. GHJ abduction				
13. External rotation @ 90°				
13. Internal rotation @ 90°				

Appendix Q. Diagnostic accuracy standardised physical assessment

"Tell me when the pain starts to come on for these tests and I will stop"
"show me where the pain is and how severe it is out of 10"(while performing the test)
Tests can be done more than once to ensure they are accurate.
If a test is painfree apply overpressure.
Resisted testing: the participant pushes against the examiner until pain or max pressure.
If bilateral pain record the subjectively "worst" shoulder

Objective assessment	Positive (provocative of their familiar/dominant pain)	Negative (not their dominant pain or less < 2/10 pain)	Unable to complete the test (why)	Indeterminate result (unclear response)
?Any evident muscle wasting of the supraspinatus, infraspinatus, pectorials, deltoid, biceps or triceps (list muscles in +ve)				
?Unilateral ACJ deformity				
?Bilateral ACJ deformity				
Spurling's test (to reproduce shoulder pain)				
Passive range of motion	Range of motion (approx.)	Reproduction of their primary pain (NPRS out of 10)	Unable to complete the test	Indeterminate result (unclear response)
Flexion Supine				
GHJ abduction in sitting (scapula fixed)				
External rotation @ 90° Supine				
Internal rotation @ 90° Supine				

Frozen shoulder is defined as 2 out of 3 or more of the following: restriction of active and passive motion of 100degrees of elevation or less, less than 50% of external rotation (at 90degrees abduction) as compared to the contralateral shoulder, and less than 50% of internal rotation. Along with absence of significant GHJ OA on xray (3 or more on the Kellgren Lawrence scale).

	Positive (provocative of their familiar/dominant pain)	Negative (not their dominant pain or less < 2/10 pain)	Pain response (NPRS out of 10)	Unable to complete the test (why)	Indeterminate result (unclear response)
1.Patient generated provocative movement: Can you show me the most painful movement for your arm?	Describe the position:				
2.External rotation adduction test					
3. AC resisted extension (resisted horizontal abduction)					
4. Bell Van Reit test. Adduct to where it hurts or end of range then resist					
5. Pain with Scapula elevation (overpressure on humeral neck)					
6. Pain with Scapula depression (overpressure on humeral neck)					

Shoulder active range of motion (affected arm only)	Range of motion (approx.)	Reproduction of their primary pain (NPRS out of 10)	Unable to complete the test	Indeterminate result (unclear response)
7. Flexion				
7. Scaption				Painful arc Y/N Pseudoparalysis Y/N
7. External rotation				
7. Internal rotation (the level of spinous process the participant can reach up their back)				

Resisted testing (compare to other side, resisted flexion/abduction push above elbow)	Strength out of 5 (Oxford Scale)	Reproduction of their primary pain (NPRS out of 10)	Unable to complete the test	Indeterminate result (unclear response)
8. Flexion 10 degrees				
8. Abduction 10 degrees				
8. External rotation (hand at side in neutral and the clinician's hand between the elbow and their side)				
8. Belly Press				

Provisional Diagnosis: .....

Post injection testing:

	Positive (provocative of their familiar/dominant pain)	Negative (not their dominant pain or less < 2/10 pain)	Pain response (NPRS out of 10)	Unable to complete the test (why)	Indeterminate result (unclear response)
1.Patient generated provocative movement					
2. Most painful test (or first most painful)					
3. 2nd most painful test					

Test	Pre injection	Post injection	% change
1 Patient generated			
2			
3			

Total % change: _____

Appendix R. Systematic review search strategies

Table 1

Reliability Search Strategy EBSCO Health Databases (MEDLINE, SPORTS Discus and CINAHL)

Search (Title + Abstract + Keywords/Subject Headings)	**Keywords (limits: 1990- 2022, English language)**
S1 (n=154,237)	(acromioclavicular OR ACJ OR "AC joint" OR subacromial OR "rotator cuff" OR "shoulder impingement" OR supraspinatus OR infraspinatus OR subscapularis OR "teres minor" OR shoulder) N5 (test* OR exam* OR measure*)
S2 (n=646,694)	(valid* OR reliable OR reliability OR inter-rater OR intra-rater) N5 (test* OR exam* OR measure*)
S3 (n=6962)	S1 and S2
S4 (n=394,690)	Kappa OR PABAK OR ICC OR "percentage agreement" OR ANOVA
S5 (n=1,440)	S3 and S4

Note. S: search. n: number. ACJ: acromioclavicular joint. N5: near operator five finds words a maximum of five words apart from one another.

Table 2

Diagnostic Accuracy Search Strategy EBSCO Health Databases (MEDLINE, SPORTS Discus and CINAHL)

Search (Title + Abstract + Keywords/Subject Headings)	**Keywords (limits: 1990- 2022, English language)**
S1 (n=8,577,671)	test* OR exam* OR "clinical test"
S2 (n=154,237)	acromioclavicular OR ACJ OR "AC joint" OR subacromial OR "rotator cuff" OR "shoulder impingement" OR supraspinatus OR infraspinatus OR subscapularis OR "teres minor" OR shoulder
S3 (n=6,032)	(shoulder) N5 (diagnos* OR accura*)
S4: (n=1,183)	S2 AND S3
S5 (n=1,052)	S1 AND S4
S6 (n=5,506,523)	specific* OR sensitiv* OR "likelihood ratio" OR "predictive value*"
S7 (n=894)	S5 AND S6

Note. S: search. n: number. ACJ: acromioclavicular joint. N5: near operator five finds words a maximum of five words apart from one another.

Appendix S. Systematic review search engines

Table 3

Search strategy for PEDro

Search	Keywords
S1: (n=964)	Abstract & Title: shoulder Therapy: Problem: pain Body Part: upper arm, shoulder or shoulder girdle Subdiscipline: musculoskeletal
S2: (n=20)	acromioclavicular
S3: (n=5) Abstract & Title	"diagnostic accuracy"
S4: (n=5)	Abstract & Title: reliability Therapy: Problem: pain Body Part: upper arm, shoulder or shoulder girdle Subdiscipline: musculoskeletal

Note. S: search. n: number.

Table 4

Reliability search strategy for AMED

Search	Keywords
S1 (n=353) abstract	reliab* AND shoulder
S2 (n=22668) abstract	Test
S3 (n=166)	S1 AND S2

Note. S: search. n: number.

Table 5

Diagnostic accuracy search strategy AMED

Search	Keywords
S1 (n= 16631) abstract	diagnos*
S2 (n=5831) abstract	accura*
S3 (n=4071) abstract	Shoulder
S4 (n=22668) abstract	test
S5 (n=101)	S1 OR S2 AND S3 AND S4

Note. S: search. n: number.

Table 6

Search strategy Cochrane Library via Wiley

	Search	Keywords
Diagnostic accuracy	search: (408) Title/abstract/keyword	acromioclavicular OR ACJ OR subacromial OR "rotator cuff" OR "shoulder impingement" OR supraspinatus OR infraspinatus OR subscapularis OR "teres minor" OR shoulder AND diagnos* OR accura* AND specific* OR sensitiv* OR "likelihood ratio" OR "predictive value*" AND test* OR exam* OR "clinical test"
Reliability	search: (90) Title/abstract/keyword	acromioclavicular OR ACJ OR subacromial OR "rotator cuff" OR "shoulder impingement" OR supraspinatus OR infraspinatus OR subscapularis OR "teres minor" OR shoulder AND valid* OR reliable OR reliability OR inter-rater OR intra-rater AND kappa OR PABAK OR ICC OR "percentage agreement" OR ANOVA AND test* OR exam* OR "clinical test"

Note. ACJ: acromioclavicular joint. PABAK: prevalence-adjusted and bias-adjusted kappa. ICC: intraclass correlation coefficient. ANOVA: analysis of variance.

Table 7

Reliability Search Strategy Scopus

Search	Keywords
S1: 933	TITLE ((acromioclavicular OR ACJ OR subacromial OR "rotator cuff" OR "shoulder impingement" OR supraspinatus OR infraspinatus OR subscapularis OR "teres minor" OR shoulder) AND (valid* OR reliable OR reliability OR inter-rater OR intra-rater))
S2: 943,687	kappa OR PABAK OR ICC OR "percentage agreement" OR ANOVA
S4: 27,912,109	test* OR exam* OR "clinical test"
S5 (306)	#1 AND #2 AND #3

Note. S: search. ACJ: acromioclavicular joint. PABAK: prevalence-adjusted and bias-adjusted kappa. ICC: intraclass correlation coefficient. ANOVA: analysis of variance.

Table 8

Diagnostic Accuracy Search Strategy Scopus

Search	Keywords
S1: 1,824	TITLE ((acromioclavicular OR ACJ OR subacromial OR "rotator cuff" OR "shoulder impingement" OR supraspinatus OR infraspinatus OR subscapularis OR "teres minor" OR shoulder) AND (diagnos* OR accura*))
S3: 20,074,073	specific* OR sensitiv* OR "likelihood ratio" OR "predictive value*"
S4: 27,912,109	test* OR exam* OR "clinical test"
S5 (580)	#1 AND #2 AND #3

Note. S: search. ACJ: acromioclavicular joint.

Appendix T. Systematic review study characteristics

	Design	No. of participants	Mean age	% Females	Clinical tests	Reference test	Inclusion criteria	Exclusion criteria
Apeldoorn et al., 2021	Inter-rater reliability study	113	38	55	Resisted ER, empty can, full can, active compression test, Neer Walsh, Hawkins-Kennedy, Resisted IR, Acromioclavicular joint stress	N/A	shoulder pain, age over 17, Dutch speaking	recent surgery, shoulder fracture, cervical nerve root entrapment total cuff tear, serious diseases (e.g., malignancies), rheumatic and/or neurological disorders, organ pathologies which affect the shoulder, dementia, and/or psychological disorders
Burns et al., 2016	Single-group repeated-measures interrater reliability study	21	41	67	Cervical and shoulder AROM & PROM. Resisted shoulder and elbow movements. Cervical spine, scapular and shoulder muscle length. Scapulo-thoracic posture. Cervical and thoracic segmental mobility	N/A	Shoulder pain, age 18-64, SPADI > 20%	Non musculoskeletal signs or symptoms, fracture, severe trauma in last 6 weeks, cervical spinal stenosis, osteoporosis, bilateral upper limb symptoms, previous cervical or thoracic spine surgery, neurological deficit, insufficient English skills.

Doxey et al., 2018	Cross-sectional inter-rater reliability study	1070	42	66	Yergason, resisted ER, Neer Walsh, painful arc, Jobe	N/A	A convenience sample of workers from 15 different companies. Workers were recruited regardless of the presence of symptoms	NR
Ferenczi et al., 2018	Cross-sectional descriptive intra and inter-rater reliability study	34	60	26	Neer Walsh and Hawkins Kennedy	N/A	Age > 40 years; shoulder pain duration > 1 month; normal passive range of motion; one or more positive clinical test results for rotator cuff testing or subacromial impingement detection; and absence of glenohumeral arthropathy or calcification on X-ray imaging.	Previous shoulder surgery, shoulder instability, humeral fracture, local steroid injections in the preceding 30 days, inflammatory joint disease, neoplastic disorders, evidence of a painful acromioclavicular joint on palpation, neurological or cervical disease
Schmidt et al., 2021	Inter-rater reliability and diagnostic accuracy study	120	56	46	Belly press, ER lag sign, IR lag sign, painful arc, Hawkins Kennedy, resisted ER	N/A	Acute soft tissue shoulder injury or successfully reduced glenohumeral dislocation and no fracture on x-ray	Injury of both shoulders, previous shoulder surgery within 6 months, rotator cuff tear on imaging, ongoing neck–/shoulder problems and other serious disease.
Lange et al., 2017	Reliability systematic review and meta-analysis	18 studies	NR	NR	Empty can, Neer Walsh, Hawkins Kennedy, painful arc	N/A	Studies assessing the intra-rater and/or inter-rater reliability of specific physical examination tests for the diagnosis of shoulder pathologies applied as a single test or in combination with other tests were included if written in English or German	Did not name or describe the physical tests if individual tests were not specified/named if only asymptomatic patients were evaluated or if the physical examination test was performed under anaesthesia or immediately postoperative. Animal studies and cadaveric

								studies and studies which used device supported testing procedures
Anauate Nicolao et al., 2022	Prospective multicentre diagnostic accuracy study	199	47	42	Painful arc, full can, empty can, Hawkins Kennedy, resisted ER, Neer Walsh, drop arm	Arthroscopy	Patients who had an indication of arthroscopy for rotator cuff tears with shoulder pain ≥4 weeks.	Adhesive capsulitis, glenohumeral osteoarthritis, or shoulder instability
Ackmann et al., 2021	Prospective single-centre diagnostic accuracy study	61	60	43	Whipple test, empty can, full can	Arthroscopy	Patients presenting for arthroscopic surgery because of rotator cuff tears	Patients with any history of shoulder surgery or adhesive capsulitis from the study
Balevi Batur et al., 2022	Prospective diagnostic accuracy study	106	55	68	Empty can, Neer Walsh, drop arm, Hawkins Kennedy, full can	MRI	Patients aged 30-65 years with shoulder pain and limited range of motion for at least four weeks	Recent history of progressive degenerative changes, shoulder surgery of fractures, malignancy, adhesive capsulitis, rotator cuff, pathologies other than supraspinatus, and infectious and inflammatory conditions of the shoulder
Yazigi et al., 2021	Prospective, multicentre diagnostic accuracy study	575	51	48	Empty-can, full can, drop-arm test, Hawkins Kennedy, Neer Walsh, painful arc, resisted ER	MRI	Patients aged ≥18 years with shoulder pain for ≥4 weeks	Adhesive capsulitis; history of fracture, dislocation, or surgery on the same limb; neck pain; substantial glenohumeral osteoarthritis with decreased joint space or subluxation seen on imaging examinations; claustrophobia; cognitive impairment of the patient that

								would prevent understanding the proposed clinical tests; missing outpatient visits; imaging examination >3 months after the patient's clinical evaluation; and traumatic or unexpected event involving the affected limb
Lädermann, Collin, et al., 2021	Diagnostic accuracy systematic review and meta-analysis	13 studies, 1699 participants	54	NR	Bear hug, belly press, IR lag sign, lift off test	Arthroscopic, MRI/MRA and ultrasound	To report at least 1 of the following: (1) true and false positives and true and false negatives; (2) sensitivity and specificity; and/or (3) positive predictive value and negative predictive value of individual clinical tests against radiographic, arthroscopic, or intraoperative observations. Diagnoses had to include rotator cuff tears involving the subscapularis. Patients had to present with shoulder pain, functional impairment, or other evidence of rotator cuff disease	Patients with shoulder injury <6 weeks, history of shoulder instability, dislocation, rheumatoid arthritis, fracture, fibromyalgia, labral lesion, adhesive capsulitis, tumor, complex regional pain syndrome, or stroke-related disorder. Articles written in languages other than English, French, German, Spanish, or Italian
Lädermann, Meynard, et al., 2021	Diagnostic accuracy systematic review and meta-analysis	14 studies, 2457 participants	54	NR	Painful arc, empty can, drop arm, Neer Walsh, Hawkins Kennedy, resisted ER, ER lag sign	Arthroscopic, MRI/MRA and ultrasound	Original clinical studies reporting true and false positives and true and false negatives AND/OR sensitivity and specificity AND/OR positive and negative predictive values of individual clinical tests, against radiographic,	Patients with history of shoulder instability, dislocations, rheumatoid arthritis, fractures, fibromyalgia, labral lesions, adhesive capsulitis, tumours, complex regional pain syndrome and stroke-related

							arthroscopic or intra-operative observations, to diagnose the presence of posterosuperior rotator cuff tears (partial or full thickness) involving the infraspinatus and/or the teres minor and/or the supraspinatus, in patients presenting with shoulder pain and/or functional impairment and/or other evidence of rotator cuff disease.	disorders. Studies reporting diagnostic accuracy of clinical tests for subacromial impingement syndrome. –Articles written in languages other than English, French, German, Italian or Spanish.
Gismervik et al., 2017	Diagnostic accuracy systematic review and meta-analysis	11 articles, 4,125 participants	47	32	Neer Walsh, Hawkins Kennedy	Arthroscopic, MRI/MRA, CT, anaesthetic injection and ultrasound	Single PETS were studied, PETS were compared to a reference test, living humans were studied, study not about fractures, dislocations of joints or nerve dysfunction, article in English or Scandinavian languages, the study included at least 20 patients, sensitivity or specificity was reported, the reference test was plausible for the condition studied, risk of bias was acceptable, requirement for pooling of data, construction of 2 × 2 contingency tables was possible	
Krill et al., 2018	Diagnostic accuracy systematic review	2 studies, 191 participants	NR	NR	Paxinos, ACJ palpation, active compression test, cross body adduction and Hawkins Kennedy	Anaesthetic injection	Level of evidence I and II studies published in peer-reviewed scientific journals focused on the physical examination, specifically evaluation of the AC joint	Non-English, not available in full-text, level of evidence III or lower, did not evaluate pathology of the AC joint, or did not apply a validated 'gold standard.'

Note. PETs: physical examination tests. AROM: active range of motion. PROM: passive range of motion. SPADI: the shoulder pain and disability index. NR: not recorded. N/A: not applicable. ER: external rotation. ACJ: acromioclavicular joint. MRI: magnetic resonance imaging. MRA: magnetic resonance arthrography.

Appendix U. Predictors not retained by the LASSO

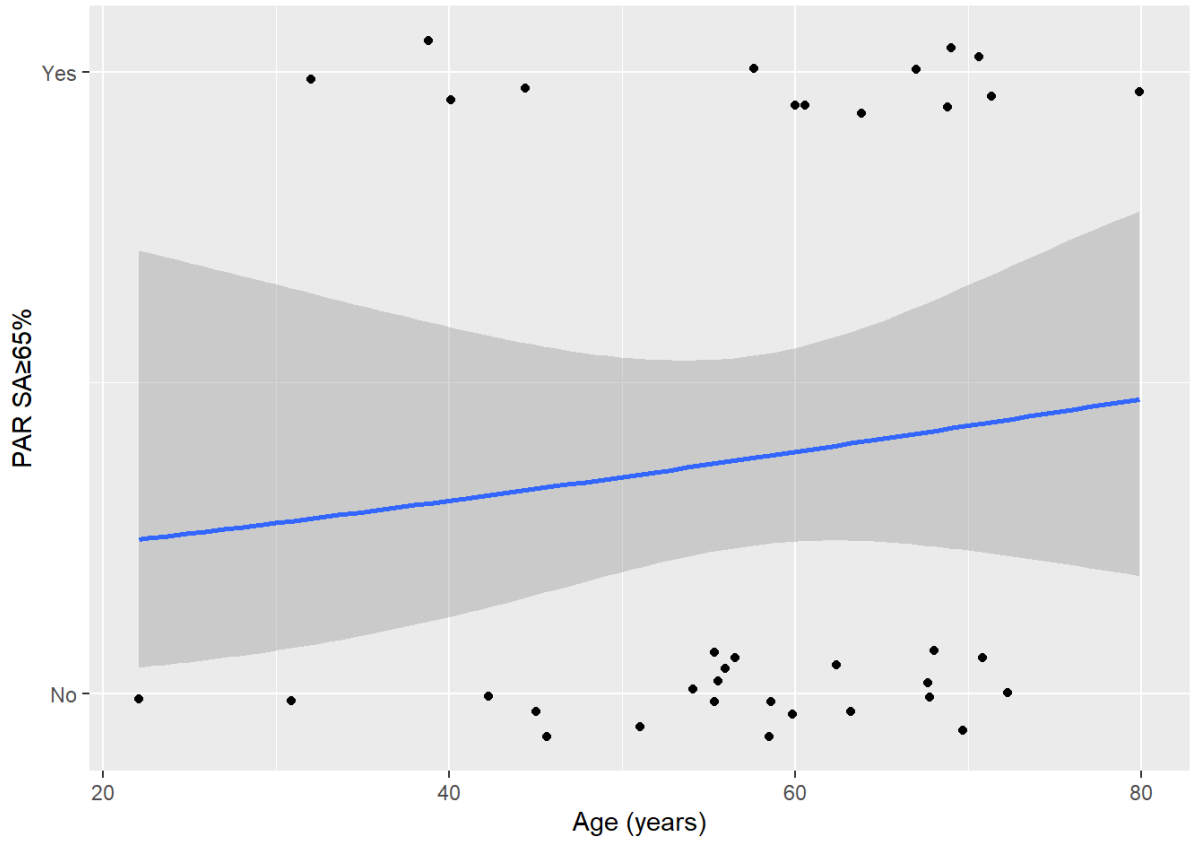
Variable	TP	FP	TN	FN	Co-efficient (95% CI)	P-value
ACJ deformity bilateral	4	4	10	19	0.62 (-0.91, 2.20)	0.42
ACJ deformity unilateral	4	5	10	18	0.37 (-1.10, 1.10)	0.62
Bell Van Riet	13	19	1	4	0.83 (-1.00, 3.30)	0.40
Resisted ACJ extension	9	15	5	8	0.08 (-1.30, 1.50)	0.91
Scapula depression	3	7	11	16	-0.40 (-2.00, 1.00)	0.59
Scapula elevation	4	8	10	15	-0.25 (-1.70, 1.10)	0.73
Belly Press	5	12	9	11	-0.63 (-2.00, 0.67)	0.35
Resisted ER	6	14	8	9	-0.63 (-2.00, 0.69)	0.35
Resisted abduction	5	13	9	10	-0.80 (-2.20, 0.50)	0.23
Passive ER	9	19	5	4	-0.92 (-2.40, 0.55)	0.22
Passive abduction	12	16	2	7	0.82 (-0.70, 2.60)	0.30
Passive flexion	12	21	2	2	-0.51 (-2.50, 1.50)	0.60
External rotation	12	22	2	1	-1.20 (-2.70, 0.23)	0.10
Painful arc	2	2	12	21	0.51 (-1.50, 2.50)	0.60
Scaption	12	21	2	2	-0.61 (-2.60, 1.40)	0.53
Flexion	11	20	3	3	-0.58 (-2.30, 1.10)	0.49
Difficulty with across chest	10	13	4	10	0.60 (-0.74, 2.00)	0.39
Difficult with reaching backwards	12	16	2	7	0.82 (-0.70, 2.60)	0.30
Difficulty with hand behind back	13	18	1	5	0.98 (-0.78, 3.30)	0.29
Crepitus	5	9	9	14	-0.20 (-1.60, 1.10)	0.77
Cannot sleep through the night	12	20	2	3	-0.16 (-1.90, 1.70)	0.86
Cannot lie on unaffected side	5	10	9	13	-0.30 (-1.70, 1.00)	0.66
Cannot lie on affected side	14	19	0	4	1.90 (-0.46, 6.80)	0.13
Secondary pain: neck pain	2	3	12	20	0.16 (-1.70, 1.90)	0.86
Secondary pain: into hand	0	1	13	22	-0.66 (-5.70, 2.30)	0.68
Secondary pain: below elbow	3	3	11	20	0.58 (-1.10, 2.30)	0.49
Primary pain: lateral shoulder	7	5	7	18	1.20 (-0.16, 2.70)	0.08
Primary pain: posterior shoulder	1	0	13	23	1.70 (-1.30, 6.70)	0.28

Variable	TP	FP	TN	FN	Co-efficient (95% CI)	P-value
Primary pain: anterior shoulder	3	6	11	17	-0.20 (-1.80, 1.30)	0.79
Pain worse at night	6	9	8	14	0.15 (-1.20, 1.50)	0.82
Constant pain	10	11	4	12	0.93 (-0.40, 2.40)	0.17
Duration >12m	11	15	3	8	0.59 (-0.82, 2.20)	0.42
MOI Insidious	4	5	10	18	0.37 (-1.10, 1.80)	0.62
MOI Injury	5	11	9	12	-0.46 (-1.80, 0.84)	0.49

Note. Imputed values for binary predictors are rounded. TP: true positive. FP: false positive. TN: true negative. FN: false negative. M: months. MOI: mechanism of injury. CI: confidence interval. ACJ: acromioclavicular joint. ER: external rotation.

Figure 1

Graphical displays of continuous predictors



Note. PAR: positive anaesthetic response. SA: subacromial.

Appendix V. Participant's predicted probability scores with the 10-predictor model

Table 9

Participant 1 predicted probability score

Coefficient factors		**Patient results**		
Factors	**Coefficients**	**Outcome**		**Score**
MOI = strain	0.41	No	(0)	0.00
MOI = repetitive	-0.31	No	(0)	0.00
Reported pain at worst ≥8/10	-0.39	Yes	(1)	-0.39
Reported pain at best ≥5/10	0.37	No	(0)	0.00
Reported 1° pain: above/on clavicle	-0.56	No	(0)	0.00
Pain with overhead tasks	0.57	Yes	(1)	0.57
Painful passive internal rotation	-0.42	Yes	(1)	-0.42
Painful adduction in external rotation	-0.68	No	(0)	0.00
Painful resisted flexion at 10^0	-0.45	No	(0)	0.00
Presence of muscle wasting	1.64	No	(0)	0.00
Linear score				-0.24
Predicted probability score				44%

Note: MOI: mechanism of injury. 1° pain: primary pain source.

Table 10

Participant 2 predicted probability score

Coefficient factors			**Patient results**	
Factors	**Coefficients**		**Outcome**	**Score**
MOI = strain	0.41	Yes	(1)	0.41
MOI = repetitive	-0.31	No	(0)	0.00
Reported pain at worst ≥8/10	-0.39	Yes	(1)	-0.39
Reported pain at best ≥5/10	0.37	Yes	(1)	0.37
Reported 1° pain: above/on clavicle	-0.56	Yes	(1)	-0.56
Pain with overhead tasks	0.57	Yes	(1)	0.57
Painful passive internal rotation	-0.42	Yes	(1)	-0.42
Painful adduction in external rotation	-0.68	Yes	(1)	-0.68
Painful resisted flexion at 10^0	-0.45	No	(0)	0.00
Presence of muscle wasting	1.64	No	(0)	0.00
Linear score				-0.69
Predicted probability score				33%

Note: MOI: mechanism of injury. 1° pain: primary pain source.

Table 11

Participant 3 predicted probability score

Coefficient factors		**Patient results**		
Factors	**Coefficients**	**Outcome**		**Score**
MOI = strain	0.41	No	(0)	0.00
MOI = repetitive	-0.31	No	(0)	0.00
Reported pain at worst ≥8/10	-0.39	Yes	(1)	-0.39
Reported pain at best ≥5/10	0.37	No	(0)	0.00
Reported 1° pain: above/on clavicle	-0.56	No	(0)	0.00
Pain with overhead tasks	0.57	Yes	(1)	0.57
Painful passive internal rotation	-0.42	Yes	(1)	-0.42
Painful adduction in external rotation	-0.68	Yes	(0)	0.00
Painful resisted flexion at 10^0	-0.45	No	(0)	0.00
Presence of muscle wasting	1.64	No	(0)	0.00
Linear score				-0.24
Predicted probability score				44%

Note: MOI: mechanism of injury. 1° pain: primary pain source.

Table 12

Participant 4 predicted probability score

Coefficient factors		**Patient results**		
Factors	**Coefficients**	**Outcome**		**Score**
MOI = strain	0.41	No	(0)	0.00
MOI = repetitive	-0.31	No	(0)	0.00
Reported pain at worst ≥8/10	-0.39	Yes	(1)	-0.39
Reported pain at best ≥5/10	0.37	No	(0)	0.00
Reported 1° pain: above/on clavicle	-0.56	No	(0)	0.00
Pain with overhead tasks	0.57	Yes	(1)	0.57
Painful passive internal rotation	-0.42	Yes	(1)	-0.42
Painful adduction in external rotation	-0.68	No	(0)	0.00
Painful resisted flexion at 10^0	-0.45	No	(0)	0.00
Presence of muscle wasting	1.64	No	(0)	0.00
Linear score				0.24
Predicted probability score				44%

Note: MOI: mechanism of injury. 1° pain: primary pain source.

Table 13

Participant 5 predicted probability score

Coefficient factors		**Patient results**		
Factors	**Coefficients**	**Outcome**		**Score**
MOI = strain	0.41	No	(0)	0.00
MOI = repetitive	-0.31	No	(0)	0.00
Reported pain at worst ≥8/10	-0.39	Yes	(1)	-0.39
Reported pain at best ≥5/10	0.37	No	(0)	0.00
Reported 1° pain: above/on clavicle	-0.56	Yes	(1)	-0.56
Pain with overhead tasks	0.57	Yes	(1)	0.57
Painful passive internal rotation	-0.42	Yes	(1)	-0.42
Painful adduction in external rotation	-0.68	No	(0)	0.00
Painful resisted flexion at 10^0	-0.45	Yes	(1)	-0.45
Presence of muscle wasting	1.64	No	(0)	0.00
Linear score				-1.24
Predicted probability score				22%

Note: MOI: mechanism of injury. 1° pain: primary pain source.

Table 14

Participant 6 predicted probability score

Coefficient factors		**Patient results**		
Factors	**Coefficients**	**Outcome**		**Score**
MOI = strain	0.41	Yes	(1)	0.41
MOI = repetitive	-0.31	No	(0)	0.00
Reported pain at worst ≥8/10	-0.39	Yes	(1)	-0.39
Reported pain at best ≥5/10	0.37	No	(0)	0.00
Reported 1° pain: above/on clavicle	-0.56	No	(0)	0.00
Pain with overhead tasks	0.57	Yes	(1)	0.57
Painful passive internal rotation	-0.42	Yes	(1)	-0.42
Painful adduction in external rotation	-0.68	Yes	(1)	-0.68
Painful resisted flexion at 10^0	-0.45	No	(0)	0.00
Presence of muscle wasting	1.64	No	(0)	0.00
Linear score				-0.51
Predicted probability score				37%

Note: MOI: mechanism of injury. 1° pain: primary pain source.

Table 15

Participant 7 predicted probability score

Coefficient factors		**Patient results**		
Factors	**Coefficients**	**Outcome**		**Score**
MOI = strain	0.41	No	(0)	0.00
MOI = repetitive	-0.31	Yes	(1)	-0.31
Reported pain at worst ≥8/10	-0.39	Yes	(1)	-0.39
Reported pain at best ≥5/10	0.37	No	(0)	0.00
Reported 1° pain: above/on clavicle	-0.56	No	(0)	0.00
Pain with overhead tasks	0.57	Yes	(1)	0.57
Painful passive internal rotation	-0.42	Yes	(1)	-0.42
Painful adduction in external rotation	-0.68	Yes	(1)	-0.68
Painful resisted flexion at 10^0	-0.45	No	(0)	0.00
Presence of muscle wasting	1.64	No	(0)	0.00
Linear score				-1.24
Predicted probability score				23%

Note: MOI: mechanism of injury. 1° pain: primary pain source.

Table 16

Participant 8 predicted probability score

Coefficient factors		**Patient results**		
Factors	**Coefficients**	**Outcome**		**Score**
MOI = strain	0.41	No	(0)	0.00
MOI = repetitive	-0.31	No	(0)	0.00
Reported pain at worst ≥8/10	-0.39	No	(0)	0.00
Reported pain at best ≥5/10	0.37	No	(0)	0.00
Reported 1° pain: above/on clavicle	-0.56	Yes	(1)	-0.56
Pain with overhead tasks	0.57	Yes	(1)	0.57
Painful passive internal rotation	-0.42	No	(0)	0.00
Painful adduction in external rotation	-0.68	Yes	(1)	-0.68
Painful resisted flexion at 10^0	-0.45	Yes	(1)	-0.45
Presence of muscle wasting	1.64	No	(0)	0.00
Linear score				-1.12
Predicted probability score				25%

Note: MOI: mechanism of injury. 1° pain: primary pain source.

Table 17

Participant 9 predicted probability score

Coefficient factors			**Patient results**	
Factors	**Coefficients**		**Outcome**	**Score**
MOI = strain	0.41	No	(0)	0.00
MOI = repetitive	-0.31	No	(0)	0.00
Reported pain at worst ≥8/10	-0.39	No	(0)	0.00
Reported pain at best ≥5/10	0.37	Yes	(1)	0.37
Reported 1° pain: above/on clavicle	-0.56	No	(0)	0.00
Pain with overhead tasks	0.57	Yes	(1)	0.57
Painful passive internal rotation	-0.42	Yes	(1)	-0.42
Painful adduction in external rotation	-0.68	Yes	(1)	-0.68
Painful resisted flexion at 10^0	-0.45	No	(0)	0.00
Presence of muscle wasting	1.64	No	(0)	0.00
Linear score				-0.16
Predicted probability score				46%

Note: MOI: mechanism of injury. 1° pain: primary pain source.

Table 18

Participant 10 predicted probability score

Coefficient factors			**Patient results**	
Factors	**Coefficients**		**Outcome**	**Score**
MOI = strain	0.41	No	(0)	0.00
MOI = repetitive	-0.31	No	(0)	0.00
Reported pain at worst ≥8/10	-0.39	Yes	(1)	-0.39
Reported pain at best ≥5/10	0.37	No	(0)	0.00
Reported 1° pain: above/on clavicle	-0.56	Yes	(1)	-0.56
Pain with overhead tasks	0.57	Yes	(1)	0.57
Painful passive internal rotation	-0.42	Yes	(1)	-0.42
Painful adduction in external rotation	-0.68	Yes	(1)	-0.68
Painful resisted flexion at 10^0	-0.45	Yes	(1)	-0.45
Presence of muscle wasting	1.64	No	(0)	0.00
Linear score				-1.93
Predicted probability score				13%

Note: MOI: mechanism of injury. 1° pain: primary pain source.

Table 19

Participant 11 predicted probability score

Coefficient factors		**Patient results**		
Factors	**Coefficients**		**Outcome**	**Score**
MOI = strain	0.41	No	(0)	0.00
MOI = repetitive	-0.31	No	(0)	0.00
Reported pain at worst ≥8/10	-0.39	Yes	(1)	-0.39
Reported pain at best ≥5/10	0.37	No	(0)	0.00
Reported 1° pain: above/on clavicle	-0.56	No	(0)	0.00
Pain with overhead tasks	0.57	Yes	(1)	0.57
Painful passive internal rotation	-0.42	Yes	(1)	-0.42
Painful adduction in external rotation	-0.68	No	(0)	0.00
Painful resisted flexion at 10^0	-0.45	Yes	(1)	-0.45
Presence of muscle wasting	1.64	No	(0)	0.00
Linear score				-0.69
Predicted probability score				33%

Note: MOI: mechanism of injury. 1° pain: primary pain source.

Table 20

Participant 12 predicted probability score

Co-efficient factors		**Patient results**		
Factors	**Coefficients**		**Outcome**	**Score**
MOI = strain	0.41	Yes	(1)	0.41
MOI = repetitive	-0.31	No	(0)	0.00
Reported pain at worst ≥8/10	-0.39	Yes	(1)	-0.39
Reported pain at best ≥5/10	0.37	No	(0)	0.00
Reported 1° pain: above/on clavicle	-0.56	No	(0)	0.00
Pain with overhead tasks	0.57	Yes	(1)	0.57
Painful passive internal rotation	-0.42	No	(0)	0.00
Painful adduction in external rotation	-0.68	Yes	(1)	-0.68
Painful resisted flexion at 10^0	-0.45	No	(0)	0.00
Presence of muscle wasting	1.64	No	(0)	0.00
Linear score				-0.09
Predicted probability score				48%

Note: MOI: mechanism of injury. 1° pain: primary pain source.

Table 21

Participant 13 predicted probability score

Coefficient factors		**Patient results**		
Factors	**Coefficients**	**Outcome**		**Score**
MOI = strain	0.41	No	(0)	0.00
MOI = repetitive	-0.31	No	(0)	0.00
Reported pain at worst ≥8/10	-0.39	Yes	(1)	-0.39
Reported pain at best ≥5/10	0.37	Yes	(1)	0.37
Reported 1° pain: above/on clavicle	-0.56	Yes	(1)	-0.56
Pain with overhead tasks	0.57	Yes	(1)	0.57
Painful passive internal rotation	-0.42	No	(0)	0.00
Painful adduction in external rotation	-0.68	Yes	(1)	-0.68
Painful resisted flexion at 10^0	-0.45	No	(0)	0.00
Presence of muscle wasting	1.64	No	(0)	0.00
Linear score				-0.69
Predicted probability score				33%

Note: MOI: mechanism of injury. 1° pain: primary pain source.

Table 22

Participant 14 predicted probability score

Coefficient factors		**Patient results**		
Factors	**Coefficients**	**Outcome**		**Score**
MOI = strain	0.41	No	(0)	0.00
MOI = repetitive	-0.31	No	(0)	0.00
Reported pain at worst ≥8/10	-0.39	Yes	(1)	-0.39
Reported pain at best ≥5/10	0.37	No	(0)	0.00
Reported 1° pain: above/on clavicle	-0.56	No	(0)	0.00
Pain with overhead tasks	0.57	Yes	(1)	0.57
Painful passive internal rotation	-0.42	Yes	(1)	-0.42
Painful adduction in external rotation	-0.68	No	(0)	0.00
Painful resisted flexion at 10^0	-0.45	Yes	(1)	-0.45
Presence of muscle wasting	1.64	No	(0)	0.00
Linear score				-0.69
Predicted probability score				33%

Note: MOI: mechanism of injury. 1° pain: primary pain source.

Table 23

Participant 15 predicted probability score

Coefficient factors		**Patient results**		
Factors	**Coefficients**	**Outcome**		**Score**
MOI = strain	0.41	No	(0)	0.00
MOI = repetitive	-0.31	No	(0)	0.00
Reported pain at worst ≥8/10	-0.39	Yes	(1)	-0.39
Reported pain at best ≥5/10	0.37	Yes	(1)	0.37
Reported 1° pain: above/on clavicle	-0.56	Yes	(1)	-0.56
Pain with overhead tasks	0.57	Yes	(1)	0.57
Painful passive internal rotation	-0.42	No	(0)	0.00
Painful adduction in external rotation	-0.68	Yes	(1)	-0.68
Painful resisted flexion at 10^0	-0.45	Yes	(1)	-0.45
Presence of muscle wasting	1.64	No	(0)	0.00
Linear score				-1.13
Predicted probability score				24%

Note: MOI: mechanism of injury. 1° pain: primary pain source.

Table 24

Participant 16 predicted probability score

Coefficient factors		**Patient results**		
Factors	**Coefficients**	**Outcome**		**Score**
MOI = strain	0.41	No	(0)	0.00
MOI = repetitive	-0.31	Yes	(1)	-0.31
Reported pain at worst ≥8/10	-0.39	No	(0)	0.00
Reported pain at best ≥5/10	0.37	Yes	(1)	0.37
Reported 1° pain: above/on clavicle	-0.56	Yes	(1)	-0.56
Pain with overhead tasks	0.57	Yes	(1)	0.57
Painful passive internal rotation	-0.42	Yes	(1)	-0.42
Painful adduction in external rotation	-0.68	Yes	(1)	-0.68
Painful resisted flexion at 10^0	-0.45	No	(0)	0.00
Presence of muscle wasting	1.64	No	(0)	0.00
Linear score				-0.13
Predicted probability score				26%

Note: MOI: mechanism of injury. 1° pain: primary pain source.

Table 25

Participant 17 predicted probability score

Coefficient factors		**Patient results**		
Factors	**Coefficients**	**Outcome**		**Score**
MOI = strain	0.41	No	(0)	0.00
MOI = repetitive	-0.31	No	(0)	0.00
Reported pain at worst ≥8/10	-0.39	No	(0)	0.00
Reported pain at best ≥5/10	0.37	No	(0)	0.00
Reported 1° pain: above/on clavicle	-0.56	Yes	(1)	-0.56
Pain with overhead tasks	0.57	Yes	(1)	0.57
Painful passive internal rotation	-0.42	Yes	(1)	-0.42
Painful adduction in external rotation	-0.68	Yes	(1)	-0.68
Painful resisted flexion at 10^0	-0.45	No	(0)	0.00
Presence of muscle wasting	1.64	No	(0)	0.00
Linear score				-1.09
Predicted probability score				25%

Note: MOI: mechanism of injury. 1° pain: primary pain source.

Table 26

Participant 18 predicted probability score

Coefficient factors		**Patient results**		
Factors	**Coefficients**	**Outcome**		**Score**
MOI = strain	0.41	No	(0)	0.00
MOI = repetitive	-0.31	Yes	(1)	-0.31
Reported pain at worst ≥8/10	-0.39	Yes	(1)	-0.39
Reported pain at best ≥5/10	0.37	No	(0)	0.00
Reported 1° pain: above/on clavicle	-0.56	Yes	(1)	-0.56
Pain with overhead tasks	0.57	Yes	(1)	0.57
Painful passive internal rotation	-0.42	No	(0)	0.00
Painful adduction in external rotation	-0.68	Yes	(1)	-0.68
Painful resisted flexion at 10^0	-0.45	No	(0)	0.00
Presence of muscle wasting	1.64	No	(0)	0.00
Linear score				-1.37
Predicted probability score				20%

Note: MOI: mechanism of injury. 1° pain: primary pain source.

Table 27

Participant 19 predicted probability score

Coefficient factors		**Patient results**		
Factors	**Coefficients**	**Outcome**		**Score**
MOI = strain	0.41	No	(0)	0.00
MOI = repetitive	-0.31	No	(0)	0.00
Reported pain at worst ≥8/10	-0.39	No	(0)	0.00
Reported pain at best ≥5/10	0.37	No	(0)	0.00
Reported 1° pain: above/on clavicle	-0.56	No	(0)	0.00
Pain with overhead tasks	0.57	Yes	(1)	0.57
Painful passive internal rotation	-0.42	Yes	(1)	-0.42
Painful adduction in external rotation	-0.68	No	(0)	0.00
Painful resisted flexion at 10^0	-0.45	No	(0)	0.00
Presence of muscle wasting	1.64	No	(0)	0.00
Linear score				0.15
Predicted probability score				54%

Note: MOI: mechanism of injury. 1° pain: primary pain source.

Table 28

Participant 20 predicted probability score

Coefficient factors		**Patient results**		
Factors	**Coefficients**	**Outcome**		**Score**
MOI = strain	0.41	No	(0)	0.00
MOI = repetitive	-0.31	Yes	(1)	-0.31
Reported pain at worst ≥8/10	-0.39	Yes	(1)	-0.39
Reported pain at best ≥5/10	0.37	No	(0)	0.00
Reported 1° pain: above/on clavicle	-0.56	No	(0)	0.00
Pain with overhead tasks	0.57	Yes	(1)	0.57
Painful passive internal rotation	-0.42	Yes	(1)	-0.42
Painful adduction in external rotation	-0.68	Yes	(1)	-0.68
Painful resisted flexion at 10^0	-0.45	Yes	(1)	-0.45
Presence of muscle wasting	1.64	No	(0)	0.00
Linear score				-1.68
Predicted probability score				16%

Note: MOI: mechanism of injury. 1° pain: primary pain source.

Table 29

Participant 21 predicted probability score

Coefficient factors		**Patient results**		
Factors	**Coefficients**	**Outcome**		**Score**
MOI = strain	0.41	No	(0)	0.00
MOI = repetitive	-0.31	Yes	(1)	-0.31
Reported pain at worst ≥8/10	-0.39	Yes	(1)	-0.39
Reported pain at best ≥5/10	0.37	No	(0)	0.00
Reported 1° pain: above/on clavicle	-0.56	No	(0)	0.00
Pain with overhead tasks	0.57	Yes	(1)	0.57
Painful passive internal rotation	-0.42	Yes	(1)	-0.42
Painful adduction in external rotation	-0.68	Yes	(1)	-0.68
Painful resisted flexion at 10^0	-0.45	Yes	(1)	-0.45
Presence of muscle wasting	1.64	Yes	(1)	1.64
Linear score				-0.04
Predicted probability score				49%

Note: MOI: mechanism of injury. 1° pain: primary pain source.

Table 30

Participant 22 predicted probability score

Coefficient factors		**Patient results**		
Factors	**Coefficients**	**Outcome**		**Score**
MOI = strain	0.41	No	(0)	0.00
MOI = repetitive	-0.31	No	(0)	0.00
Reported pain at worst ≥8/10	-0.39	Yes	(1)	-0.39
Reported pain at best ≥5/10	0.37	No	(0)	0.00
Reported 1° pain: above/on clavicle	-0.56	Yes	(1)	-0.56
Pain with overhead tasks	0.57	Yes	(1)	0.57
Painful passive internal rotation	-0.42	No	(0)	0.00
Painful adduction in external rotation	-0.68	Yes	(1)	-0.68
Painful resisted flexion at 10^0	-0.45	No	(0)	0.00
Presence of muscle wasting	1.64	Yes	(1)	1.64
Linear score				0.58
Predicted probability score				64%

Note: MOI: mechanism of injury. 1° pain: primary pain source.

Table 31

Participant 23 predicted probability score

Coefficient factors		**Patient results**		
Factors	**Coefficients**	**Outcome**		**Score**
MOI = strain	0.41	No	(0)	0.00
MOI = repetitive	-0.31	No	(0)	0.00
Reported pain at worst ≥8/10	-0.39	Yes	(1)	-0.39
Reported pain at best ≥5/10	0.37	No	(0)	0.00
Reported 1° pain: above/on clavicle	-0.56	Yes	(1)	-0.56
Pain with overhead tasks	0.57	No	(0)	0.00
Painful passive internal rotation	-0.42	Yes	(1)	-0.42
Painful adduction in external rotation	-0.68	No	(0)	0.00
Painful resisted flexion at 10^0	-0.45	No	(0)	0.00
Presence of muscle wasting	1.64	No	(0)	0.00
Linear score				-1.36
Predicted probability score				20%

Note: MOI: mechanism of injury. 1° pain: primary pain source.

Table 32

Participant 24 predicted probability score

Coefficient factors		**Patient results**		
Factors	**Coefficients**	**Outcome**		**Score**
MOI = strain	0.41	No	(0)	0.00
MOI = repetitive	-0.31	Yes	(1)	-0.31
Reported pain at worst ≥8/10	-0.39	No	(0)	0.00
Reported pain at best ≥5/10	0.37	No	(0)	0.00
Reported 1° pain: above/on clavicle	-0.56	No	(0)	0.00
Pain with overhead tasks	0.57	No	(0)	0.00
Painful passive internal rotation	-0.42	No	(0)	0.00
Painful adduction in external rotation	-0.68	Yes	(1)	-0.68
Painful resisted flexion at 10^0	-0.45	No	(0)	0.00
Presence of muscle wasting	1.64	No	(0)	0.00
Linear score				-0.99
Predicted probability score				27%

Note: MOI: mechanism of injury. 1° pain: primary pain source.

Table 33

Participant 25 predicted probability score

Coefficient factors		**Patient results**		
Factors	**Coefficients**	**Outcome**		**Score**
MOI = strain	0.41	Yes	(1)	0.41
MOI = repetitive	-0.31	No	(0)	0.00
Reported pain at worst ≥8/10	-0.39	Yes	(1)	-0.39
Reported pain at best ≥5/10	0.37	Yes	(1)	0.37
Reported 1° pain: above/on clavicle	-0.56	Yes	(1)	-0.56
Pain with overhead tasks	0.57	Yes	(1)	0.57
Painful passive internal rotation	-0.42	Yes	(1)	-0.42
Painful adduction in external rotation	-0.68	No	(0)	0.00
Painful resisted flexion at 10^0	-0.45	Yes	(1)	-0.45
Presence of muscle wasting	1.64	Yes	(1)	1.64
Linear score				1.17
Predicted probability score				76%

Note: MOI: mechanism of injury. 1° pain: primary pain source.

Table 34

Participant 26 predicted probability score

Co-efficient factors		**Patient results**		
Factors	**Coefficients**	**Outcome**		**Score**
MOI = strain	0.41	No	(0)	0.00
MOI = repetitive	-0.31	No	(0)	0.00
Reported pain at worst ≥8/10	-0.39	Yes	(1)	-0.39
Reported pain at best ≥5/10	0.37	No	(0)	0.00
Reported 1° pain: above/on clavicle	-0.56	Yes	(1)	-0.56
Pain with overhead tasks	0.57	Yes	(1)	0.57
Painful passive internal rotation	-0.42	Yes	(1)	-0.42
Painful adduction in external rotation	-0.68	Yes	(1)	-0.68
Painful resisted flexion at 10^0	-0.45	Yes	(1)	-0.45
Presence of muscle wasting	1.64	No	(0)	0.00
Linear score				-1.93
Predicted probability score				13%

Note: MOI: mechanism of injury. 1° pain: primary pain source.

Table 35

Participant 27 predicted probability score

Coefficient factors		**Patient results**		
Factors	**Coefficients**	**Outcome**		**Score**
MOI = strain	0.41	No	(0)	0.00
MOI = repetitive	-0.31	No	(0)	0.00
Reported pain at worst ≥8/10	-0.39	No	(0)	0.00
Reported pain at best ≥5/10	0.37	No	(0)	0.00
Reported 1° pain: above/on clavicle	-0.56	No	(0)	0.00
Pain with overhead tasks	0.57	Yes	(1)	0.57
Painful passive internal rotation	-0.42	No	(0)	0.00
Painful adduction in external rotation	-0.68	No	(0)	0.00
Painful resisted flexion at 10^0	-0.45	Yes	(1)	-0.45
Presence of muscle wasting	1.64	No	(0)	0.00
Linear score				0.12
Predicted probability score				53%

Note: MOI: mechanism of injury. 1° pain: primary pain source.

Table 36

Participant 28 predicted probability score

Coefficient factors		**Patient results**		
Factors	**Coefficients**	**Outcome**		**Score**
MOI = strain	0.41	No	(0)	0.00
MOI = repetitive	-0.31	No	(0)	0.00
Reported pain at worst ≥8/10	-0.39	Yes	(1)	-0.39
Reported pain at best ≥5/10	0.37	Yes	(1)	0.37
Reported 1° pain: above/on clavicle	-0.56	No	(0)	0.00
Pain with overhead tasks	0.57	No	(0)	0.00
Painful passive internal rotation	-0.42	Yes	(1)	-0.42
Painful adduction in external rotation	-0.68	No	(0)	0.00
Painful resisted flexion at 10^0	-0.45	Yes	(1)	-0.45
Presence of muscle wasting	1.64	No	(0)	0.00
Linear score				-0.88
Predicted probability score				29%

Note: MOI: mechanism of injury. 1° pain: primary pain source.

Table 37

Participant 29 predicted probability score

Coefficient factors		**Patient results**		
Factors	**Coefficients**	**Outcome**		**Score**
MOI = strain	0.41	No	(0)	0.000
MOI = repetitive	-0.31	No	(0)	0.000
Reported pain at worst ≥8/10	-0.39	Yes	(1)	-0.390
Reported pain at best ≥5/10	0.37	No	(0)	0.000
Reported 1° pain: above/on clavicle	-0.56	No	(0)	0.000
Pain with overhead tasks	0.57	Yes	(1)	0.566
Painful passive internal rotation	-0.42	Yes	(1)	-0.419
Painful adduction in external rotation	-0.68	Yes	(1)	-0.682
Painful resisted flexion at 10^0	-0.45	Yes	(1)	-0.445
Presence of muscle wasting	1.64	No	(0)	0.000
Linear score				-1.37
Predicted probability score				20%

Note: MOI: mechanism of injury. 1° pain: primary pain source.

Table 38

Participant 30 predicted probability score

Coefficient factors		**Patient results**		
Factors	**Coefficients**	**Outcome**		**Score**
MOI = strain	0.41	No	(0)	0.00
MOI = repetitive	-0.31	No	(0)	0.00
Reported pain at worst ≥8/10	-0.39	Yes	(1)	-0.39
Reported pain at best ≥5/10	0.37	No	(0)	0.00
Reported 1° pain: above/on clavicle	-0.56	Yes	(1)	-0.56
Pain with overhead tasks	0.57	Yes	(1)	0.57
Painful passive internal rotation	-0.42	Yes	(1)	-0.42
Painful adduction in external rotation	-0.68	Yes	(1)	-0.68
Painful resisted flexion at 10^0	-0.45	Yes	(1)	-0.45
Presence of muscle wasting	1.64	No	(0)	0.00
Linear score				-1.93
Predicted probability score				13%

Note: MOI: mechanism of injury. 1° pain: primary pain source.

Table 39

Participant 31 predicted probability score

Coefficient factors		**Patient results**		
Factors	**Coefficients**	**Outcome**		**Score**
MOI = strain	0.41	No	(0)	0.00
MOI = repetitive	-0.31	Yes	(1)	-0.31
Reported pain at worst ≥8/10	-0.39	No	(0)	0.00
Reported pain at best ≥5/10	0.37	Yes	(1)	0.37
Reported 1° pain: above/on clavicle	-0.56	No	(0)	0.00
Pain with overhead tasks	0.57	Yes	(1)	0.57
Painful passive internal rotation	-0.42	No	(0)	0.00
Painful adduction in external rotation	-0.68	No	(0)	0.00
Painful resisted flexion at 10^0	-0.45	No	(0)	0.00
Presence of muscle wasting	1.64	No	(0)	0.00
Linear score				0.63
Predicted probability score				65%

Note: MOI: mechanism of injury. 1° pain: primary pain source.

***Table* 40**

Participant 32 predicted probability score

Coefficient factors		**Patient results**		
Factors	**Coefficients**	**Outcome**		**Score**
MOI = strain	0.41	No	(0)	0.00
MOI = repetitive	-0.31	No	(0)	0.00
Reported pain at worst ≥8/10	-0.39	Yes	(1)	-0.39
Reported pain at best ≥5/10	0.37	No	(0)	0.00
Reported 1° pain: above/on clavicle	-0.56	No	(0)	0.00
Pain with overhead tasks	0.57	No	(0)	0.00
Painful passive internal rotation	-0.42	No	(0)	0.00
Painful adduction in external rotation	-0.68	Yes	(1)	-0.68
Painful resisted flexion at 10^0	-0.45	Yes	(1)	-0.45
Presence of muscle wasting	1.64	No	(0)	0.00
Linear score				-1.52
Predicted probability score				18%

Note: MOI: mechanism of injury. 1° pain: primary pain source.

Table 41

Participant 33 predicted probability score

Coefficient factors		**Patient results**		
Factors	**Coefficients**	**Outcome**		**Score**
MOI = strain	0.41	No	(0)	0.00
MOI = repetitive	-0.31	No	(0)	0.00
Reported pain at worst ≥8/10	-0.39	No	(0)	0.00
Reported pain at best ≥5/10	0.37	Yes	(1)	0.37
Reported 1° pain: above/on clavicle	-0.56	No	(0)	0.00
Pain with overhead tasks	0.57	Yes	(1)	0.57
Painful passive internal rotation	-0.42	No	(0)	0.00
Painful adduction in external rotation	-0.68	No	(0)	0.00
Painful resisted flexion at 10^0	-0.45	No	(0)	0.00
Presence of muscle wasting	1.64	No	(0)	0.00
Linear score				0.94
Predicted probability score				72%

Note: MOI: mechanism of injury. 1° pain: primary pain source.

Table 42

Participant 34 predicted probability score

Coefficient factors		**Patient results**		
Factors	**Coefficients**	**Outcome**		**Score**
MOI = strain	0.41	No	(0)	0.00
MOI = repetitive	-0.31	No	(0)	0.00
Reported pain at worst ≥8/10	-0.39	Yes	(1)	-0.39
Reported pain at best ≥5/10	0.37	No	(0)	0.00
Reported 1° pain: above/on clavicle	-0.56	No	(0)	0.00
Pain with overhead tasks	0.57	Yes	(1)	0.57
Painful passive internal rotation	-0.42	Yes	(1)	-0.42
Painful adduction in external rotation	-0.68	Yes	(1)	-0.68
Painful resisted flexion at 10^0	-0.45	Yes	(1)	-0.45
Presence of muscle wasting	1.64	No	(0)	1.64
Linear score				0.27
Predicted probability score				57%

Note: MOI: mechanism of injury. 1° pain: primary pain source.

Table 43

Participant 35 predicted probability score

Coefficient factors		**Patient results**		
Factors	**Coefficients**	**Outcome**		**Score**
MOI = strain	0.41	No	(0)	0.00
MOI = repetitive	-0.31	No	(0)	0.00
Reported pain at worst ≥8/10	-0.39	Yes	(1)	-0.39
Reported pain at best ≥5/10	0.37	No	(0)	0.00
Reported 1° pain: above/on clavicle	-0.56	No	(0)	0.00
Pain with overhead tasks	0.57	No	(0)	0.00
Painful passive internal rotation	-0.42	Yes	(1)	-0.42
Painful adduction in external rotation	-0.68	Yes	(1)	-0.68
Painful resisted flexion at 10^0	-0.45	Yes	(1)	-0.45
Presence of muscle wasting	1.64	No	(0)	0.00
Linear score				-1.94
Predicted probability score				13%

Note: MOI: mechanism of injury. 1° pain: primary pain source.

Table 44

Participant 36 predicted probability score

Coefficient factors		**Patient results**		
Factors	**Coefficients**	**Outcome**		**Score**
MOI = strain	0.41	No	(0)	0.00
MOI = repetitive	-0.31	No	(0)	0.00
Reported pain at worst ≥8/10	-0.39	No	(0)	0.00
Reported pain at best ≥5/10	0.37	No	(0)	0.00
Reported 1° pain: above/on clavicle	-0.56	No	(0)	0.00
Pain with overhead tasks	0.57	Yes	(1)	0.57
Painful passive internal rotation	-0.42	Yes	(1)	-0.42
Painful adduction in external rotation	-0.68	Yes	(1)	-0.68
Painful resisted flexion at 10^0	-0.45	Yes	(1)	-0.45
Presence of muscle wasting	1.64	Yes	(1)	1.64
Linear score				0.66
Predicted probability score				66%

Note: MOI: mechanism of injury. 1° pain: primary pain source.

Table 45

Participant 37 predicted probability score

Coefficient factors		**Patient results**		
Factors	**Coefficients**	**Outcome**		**Score**
MOI = strain	0.41	Yes	(1)	0.41
MOI = repetitive	-0.31	No	(0)	0.00
Reported pain at worst ≥8/10	-0.39	Yes	(1)	-0.39
Reported pain at best ≥5/10	0.37	Yes	(1)	0.37
Reported 1° pain: above/on clavicle	-0.56	No	(0)	0.00
Pain with overhead tasks	0.57	Yes	(1)	0.57
Painful passive internal rotation	-0.42	Yes	(1)	-0.42
Painful adduction in external rotation	-0.68	No	(0)	0.00
Painful resisted flexion at 10^0	-0.45	No	(0)	0.00
Presence of muscle wasting	1.64	Yes	(1)	1.64
Linear score				2.18
Predicted probability score				90%

Note: MOI: mechanism of injury. 1° pain: primary pain source.