

RESEARCH ARTICLE

 OPEN ACCESS

## Intertester reliability of shoulder complaints diagnoses in primary health care

Benny Storheil<sup>a</sup>, Elise Klouman<sup>b</sup>, Stian Holmvik<sup>b</sup>, Nina Emaus<sup>a</sup> and Nils Fleten<sup>b</sup>

<sup>a</sup>Department of Health and Care Sciences, UiT the Arctic University of Norway, Tromsø, Norway; <sup>b</sup>Department of Community Medicine, UiT the Arctic University of Norway, Tromsø, Norway

### ABSTRACT

**Objective:** Shoulder complaints are frequently encountered in general practice, but precise diagnosing is challenging. This study investigated agreement of shoulder complaints diagnoses between clinicians in a primary health care setting.

**Design:** Cross-sectional study.

**Setting:** Four primary health care clinicians used patients' history and functional examination of the shoulder by selective tissue tension techniques (STTs), to diagnose shoulder complaints.

**Subjects:** 62 patients, aged 18–75 years.

**Main outcome measure:** Reliability of diagnoses was assessed by observed intertester agreement and Cohen's kappa. A total of 372 diagnostic pairs were available for intertester comparisons.

**Results:** Six diagnoses were assigned by all clinicians; supraspinatus-, infraspinatus-, subscapularis-tendinopathies; chronic subacromial bursitis; glenohumeral capsulitis, and acromioclavicular joint lesion. The observed agreement on these diagnoses ranged from 0.84 for glenohumeral capsulitis to 0.97 for acromioclavicular joint lesion. Kappa scores were 0.46 (95% CI 0.33, 0.58) for chronic subacromial bursitis; 0.53 (95% CI 0.34, 0.68), 0.59 (95% CI 0.47, 0.70), and 0.68 (95% CI 0.53, 0.82) for infraspinatus -, supraspinatus -, and subscapularis-tendinopathy, respectively. For glenohumeral capsulitis and acromioclavicular lesion kappa scores were 0.66 (95% CI 0.57, 0.73) and 0.78 (95% CI 0.61, 0.90). Kappa scores were higher for individual diagnoses than for individual tests, except for limitation in passive abduction (0.70, 95% CI 0.62, 0.78) and passive lateral rotation (0.66, 95% CI 0.57, 0.73).

**Conclusions:** Although experienced clinicians showed substantial intertester agreement, precise diagnoses of shoulder complaints in primary health care remain a challenge. The present results call for further research on refined diagnoses of shoulder complaints.

### KEY POINTS

Based on medical history and a systematic functional examination by selective tissue tension techniques (STTs), we investigated the agreement of shoulder complaints diagnoses across four primary health care clinicians and 62 patients.

- Agreements on diagnoses were generally better than the agreement on individual tests.
- Good kappa scores were obtained for the diagnoses glenohumeral capsulitis, rotator cuff tendinopathy, and acromioclavicular lesion.
- Further research is necessary to investigate the diagnostic validity of functional shoulder examination by the STTs method.

### ARTICLE HISTORY

Received 28 May 2015  
Accepted 23 March 2016

### KEYWORDS

Cyriax; general practice; Norway; orthopaedic medicine; selective tissue tension techniques; shoulder complaints; shoulder diagnosis; shoulder pain; Sweden

## Introduction

The prevalence of musculoskeletal complaints in the general population is high and represents a major socio-economic burden.[1] Shoulder pain is among the most frequent [2] and costly [3] complaints presented in primary health care. Moreover, shoulder pain has little tendency to resolve quickly or completely; according to a Dutch study, one-half of all sufferers reported problems one year after their initial consultation.[4] Musculoskeletal complaints and disorders have

traditionally been a grossly neglected area of medical education.[5] The combination of inadequate education and high incidence of such complaints lead to a large number of patients for whom there is neither clear diagnosis nor proper treatment.[6] In many cases chronicity, frustration and social security expenditures are inevitable consequences.[7] Diagnoses of musculoskeletal disorders have proved difficult with objective methods including medical imaging.[8] Available clinical examinations involve a multitude of tests, and

CONTACT Nils Fleten  Nils.Fleten@uit.no  Department of Community Medicine, UiT the Arctic University of Norway, N-9037 Tromsø, Norway

© 2016 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

more than one hundred separate clinical shoulder tests are described, without clear advice on the most appropriate tests.[9,10]. Furthermore, a review by Hegedus et al. indicated that very few clinical shoulder tests appear to be diagnostically discriminatory.[11] However, Hegedus et al. only looked at the value of each individual test, and their updated review stated a clear need for valid, comprehensible, systematic shoulder diagnostics.[12] A proper diagnostic system is a prerequisite for clinical management; i.e., synonymous labelling of the same clinical conditions, thereby enabling specific treatment and evaluation of treatment outcome.[13].

Orthopaedic medicine ad modum Cyriax is a diagnostic system of the loco motor system founded on the theories of referred pain and selective tissue tension techniques (STTs).[14] Diagnosis based on thorough anamneses and functional examination consisting of standardized clinical tests that differentiate lesions of contractile and stretchable structures such as muscle, tendon, and tendon insertion from inert structures. A detailed clinical examination using these principles may outline clinical patterns. STTs, originally described by Dr Cyriax, represent a systematic approach to the physical assessment and diagnosis of musculoskeletal soft-tissue disorders with a potential to standardize the diagnostic labelling of shoulder complaints.[14] Several educational organizations worldwide offer STTs training,[15] and reliability studies of STTs have demonstrated adequate results.[14,16] The aim of this study was to explore the agreement of shoulder complaints diagnoses between experienced clinicians practicing systematic examination with STTs in primary health care.

## Material and methods

In April 2013, we conducted a cross-sectional study in the city of Bodø in Northern Norway. We recruited altogether 62 patients with shoulder complaints encountered in primary health care through referrals from general practitioners (GPs), physical therapists, and patients referred to the department of orthopedic surgery from GPs in the catchment area. Patients 18 years of age or above with shoulder pain of minimum 14 days were eligible. We excluded from participation patients with infectious disease, implants in the affected shoulder, or serious psychiatric disorders. The study was conducted over two days with clinical examinations performed by three physical therapists and one GP, all trained in STTs, according to a common protocol. To reflect ordinary practice, no

pre-calibration was undertaken. Each clinician was from a different clinic in Norway or Sweden. All participants gave written informed consent following written and verbal explanation of the study aims. Emphasis on the potential discomfort associated with repeated physical testing, and the option to withdraw was provided.

### **Medical history, clinical examination, and shoulder diagnosis**

Before clinical examination, all patients were interviewed and a detailed medical history was obtained by one of the researchers (BS), using a standardized questionnaire constructed in Questback. BS did not participate in the physical examination of patients. He recorded the responses electronically, and identical copies of responses were printed, and provided to all four clinicians. Each patient consecutively underwent clinical examinations in separate rooms by all four clinicians. To ensure standardization, all clinicians received the medical history, and they registered their findings and diagnostic conclusion(s) digitally in a standardized recording form on their personal computer.

Prior to the study, the clinicians evaluated the medical history questionnaire and the recording form, stating that it met the specifications and fulfilled the intended purpose of the diagnostic system. The questionnaire and the recording form ensured consistent protocol application in a clinically realistic environment. The first part of the recording form attempted to elucidate whether any symptoms in the arm originated from a shoulder lesion. If this was not clear from the information provided in the questionnaire, the clinicians performed a preliminary examination, including tests of the cervical spine, shoulder, and elbow.

The basic functional examination of the shoulder consists of 12 tests, Table 1.[17] It is important to perform all of these tests, and not to stop even if the diagnosis appears clear after a limited number of tests. If the diagnosis remains unclear after the basic functional examination, an accessory test can be used (passive horizontal adduction) to arrive at a final diagnosis. The four clinicians were instructed to recognize “inherent likelihoods”, a term defined as the sequence of symptoms and/or signs that belong to the clinical picture of a certain clinical entity that are likely to be found, in a sequence that is more or less typical for that clinical picture. The clinicians received a list of possible diagnoses they could assign, allowing combinations, to ensure use of the same diagnostic nomenclature. The clinicians were blinded to each other’s results during the clinical examinations, and patients

instructed not to communicate any specific information about previous clinical examinations to any clinician. The patients got a 5–10 min break after each clinical examination. A member of the research team observed the clinical examinations (NE).

### Statistics

The recorded information was made available in excel and converted to a SPSS file. In the absence of a gold standard for the possible diagnoses, diagnostic reliability was analysed by intertester agreement as observed agreement (Po) and agreement beyond chance, Cohen's kappa.[18] According to Altman (p. 404), a kappa value ranging between 41 and 60 was interpreted as moderate agreement and values from 61 to 80 as good agreement.[19] A 95% confidence interval (CI) for kappa was calculated by the bootstrapping

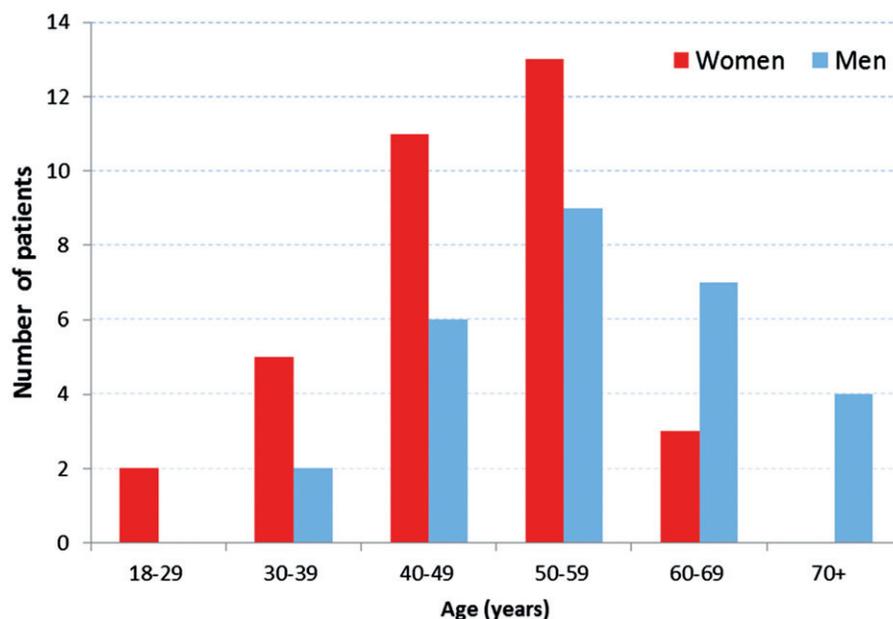
**Table 1.** The basic functional shoulder tests in selective tissue tension techniques (STTs), to diagnose shoulder complaints.

Summary of the basic functional examination of the shoulder	
Elevation	1. Active elevation of both arms. 2. Passive elevation 3. Painful arc
Glenohumeral joint	4. Passive scapulohumeral abduction 5. Passive lateral rotation 6. Passive medial rotation
Resisted movements	7. Resisted adduction 8. Resisted abduction 9. Resisted lateral rotation 10. Resisted medial rotation 11. Resisted flexion of the elbow 12. Resisted extension of the elbow
Additional test allowed in this study.	13. Passive horizontal adduction

procedure in SPSS version 19 (SPSS Inc., Chicago, IL) with samples set to 1000.[20] Power calculation indicated some 300 comparisons to identify a 0.2 difference in kappa value at 0.8 probability and significance level 0.05 for the more prevalent shoulder diagnoses. For the agreement analyses, we rearranged the data to contain six pairs (AB, AC, AD, BC, BD, and CD) of clinicians for each patient, giving a total of 372 comparisons of agreements of the 62 patients. Intertester agreement was determined at three diagnostic levels: individual tests, individual diagnoses (glenohumeral capsulitis, acromioclavicular joint lesion), and diagnostic groups: rotator cuff tendinopathies (supraspinatus-, infraspinatus-, and subscapularis-tendinopathies) and bursitis (chronic and acute subacromial-, and subdeltoid bursitis).

### Results

The study cohort included 34 women and 28 men between 18 and 75 years of age, of whom 63% were between 40 and 60 years of age (Figure 1). Women dominated in the age groups of 50 years and under, and men dominated in the age groups of 60 years or over. More than 90% of patients reported localization of symptoms to the shoulder, 89% to upper arm, 45% to the forearm and 27% to the fingers. In addition to pain (100%), 23 and 24% of patients reported paresthesia and numbness, respectively. In the medical history, 95% of participants reported symptoms to the C5 dermatome. Patient-reported causes of symptoms were equally distributed between overuse, trauma, and



**Figure 1.** Age distribution of the patients included in the study by sex (34 women and 28 men).

spontaneously occurring. In 75% of patients, the symptoms had lasted more than 26 weeks, and in 23%, 8–26 weeks. Sixty-eight percent of patients reported pain at rest, 90% activity-related pain, 71% nocturnal pain, and more than 70% reported problems lying on the affected shoulder.

### Intertester agreement on individual tests performed with STT

Considering individual tests the kappa score for resisted abduction pain was 0.26, for resisted lateral rotation pain 0.40, and resisted medial rotation pain 0.23 (Table 2). Painful arc revealed kappa scores of 0.30 and 0.16 for the agreement of a positive or negative painful arc, respectively. The kappa score for limitation of passive lateral rotation and passive abduction were 0.66 and 0.70, respectively, compared to kappa 0.45 for limitation of passive medial rotation.

### Intertester agreement on individual diagnoses of shoulder complaints

Observed agreement on the necessity of a preliminary examination of the cervical spine was high. Due to three missing, 363 pairs were available for analysis, 115 pairs agreed on the necessity of preliminary examination, and 226 agreed on no necessity of this test (22 pairs disagreed), observed agreement 0.94, kappa 0.87 (95% CI 0.81, 0.95). All clinicians agreed on six specific diagnoses in at least one patient: supraspinatus-, infraspinatus-, and subscapularis-tendinopathies, chronic subacromial bursitis, glenohumeral capsulitis, and acromioclavicular joint lesion (Figure 2). At least one clinician diagnosed one-half of the patients with glenohumeral capsulitis, and all four clinicians agreed that 12 of the patients met the diagnostic criteria for this diagnosis (Figure 2). Of the 17 patients given a diagnosis of supraspinatus-tendinopathy, five received this diagnosis from all four clinicians.

**Table 2.** Observed and chance corrected agreement (kappa) and with 95% confidence interval for the individual soft tissue tension techniques (STTs) that make up the basic functional examination of the shoulder in diagnostic assessment of 62 patients with shoulder complaints in a primary health care setting.

STTs	Signs and symptoms	N of pairs agreed		Observed agreement	Kappa	95% CI Kappa
		Yes	No			
Active elevation both arms	Pain	271	25	0.80	0.27	0.15–0.40
	Weakness	134	134	0.72	0.45	0.35–0.54
	Negative	14	302	0.85	0.26	0.11–0.40
Passive elevation affected arm	Pain	262	34	0.80	0.35	0.23–0.47
	Limitation	106	169	0.74	0.46	0.37–0.56
	Negative	26	287	0.73	0.38	0.24–0.50
Painful arc	Yes	39	240	0.75	0.30	0.19–0.41
Painful arc	No	173	50	0.60	0.16	0.08–0.22
Passive abduction shoulder	Pain	120	135	0.69	0.37	0.27–0.46
	Limitation	116	203	0.86	0.70	0.62–0.78
	Negative	110	152	0.70	0.40	0.30–0.49
Passive lateral rotation shoulder	Pain	231	33	0.71	0.19	0.08–0.30
	Limitation	125	185	0.83	0.66	0.57–0.73
	Negative	27	252	0.75	0.21	0.09–0.33
Passive medial rotation shoulder	Pain	185	68	0.68	0.29	0.20–0.39
	Limitation	47	254	0.81	0.45	0.35–0.56
	Negative	56	212	0.72	0.34	0.24–0.44
Resisted adduction shoulder	Pain	4	307	0.84	0.03	n0.07–0.14
	Weakness	0	372	1.00		
	Negative	292	7	0.80	0.05	n0.06–0.16
Resisted abduction shoulder	Pain	108	126	0.63	0.26	0.17–0.36
	Weakness	0	324	0.87	n0.07	n0.08– n0.04
	Negative	111	114	0.60	0.21	0.11–0.31
Resisted lateral rotation shoulder	Pain	78	192	0.73	0.40	0.30–0.49
	Weakness	3	345	0.94	0.17	n0.03–0.37
	Negative	174	84	0.69	0.35	0.25–0.45
Resisted medial rotation shoulder	Pain	49	202	0.67	0.23	0.13–0.33
	Weakness	0	366	0.98	n0.01	n0.02–0.00
	Negative	171	57	0.61	0.15	0.04–0.25
Resisted flexion elbow	Pain	3	330	0.90	0.08	n0.05–0.22
	Weakness	0	369	0.99	0.00	n0.01–0.00
	Negative	303	3	0.90	n0.01	n0.09–0.09
Resisted extension elbow	Pain	6	315	0.86	0.12	n0.01–0.25
	Weakness	4	340	0.92	0.18	0.01–0.36
	Negative	285	18	0.81	0.24	0.11–0.37

Numbers of the 372 assessment pairs that agreed on yes or no, respectively, are presented. *n* indicates a negative value.

Observed agreement ranged from 0.84 for glenohumeral capsulitis to 0.97 for the acromioclavicular joint lesion. The corresponding kappa scores were 0.66 (95% CI 0.57, 0.73) for glenohumeral capsulitis, 0.78 (95% CI 0.61, 0.90) for acromioclavicular lesion, 0.53 (95% CI 0.34, 0.68) for infraspinatus-, 0.59 (95% CI 0.47, 0.70) for supraspinatus-, and 0.68 (95% CI 0.53, 0.82) for subscapularis-tendinopathy, respectively (Table 3).

### Intertester agreement on specific diagnostic groups

Collapsed into a single variable as rotator cuff tendinopathy, the observed agreement was 0.88 and the kappa score was 0.66 (95% CI 0.57, 0.75). The kappa scores for each pair varied between 0.60 and 0.71 for the collapsed rotator cuff tendinopathy. Considering bursitis as a single group collapsed into a single

variable as presented in Table 3, the observed agreement was 0.84 with a kappa score of 0.43 (95% CI 0.30, 0.55). The kappa score for chronic subacromial bursitis was 0.46 (95% CI 0.33, 0.58) (Table 3).

In sex-specific analyses, kappa scores were generally higher in women, but did not reach statistical significance, although the kappa scores for rotator cuff tendinopathy were 0.59 for men and 0.72 for women (data not shown).

## Discussion

### Principal finding

Experienced clinicians, trained in primary health care, carried out functional testing of the shoulder on the same patients consecutively. The study revealed excellent agreement on needs for cervical spine examination, good agreement on glenohumeral capsulitis,

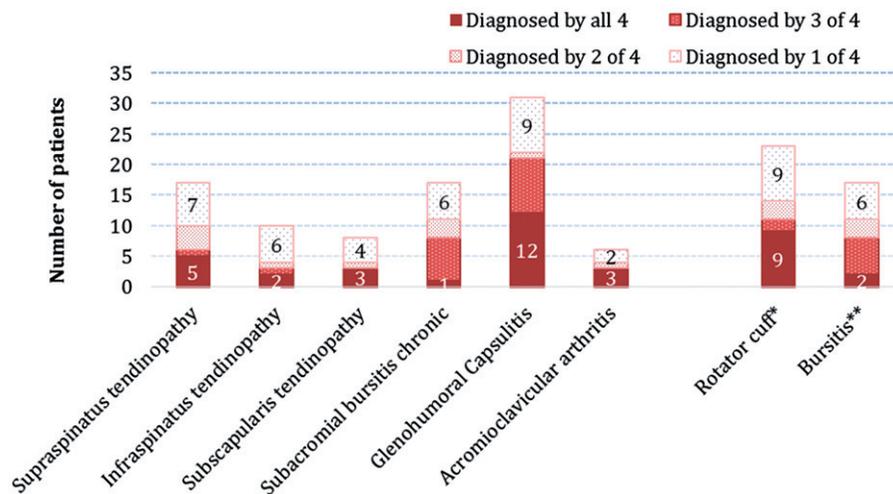


Figure 2. Number of the 62 patients given specific diagnosis by all four, three of four, two of four, or one of four of the clinicians.

\*Rotator cuff includes supraspinatus-, infraspinatus-, and subscapularis-tendinopathies;

\*\*bursitis includes chronic and acute subacromial and subdeltoid bursitis.

Table 3. Mean and range proportion of 62 patients with shoulder complaints in a primary health care setting given a specific diagnosis by four independent clinicians.

	Mean proportion	Range proportion	N of pairs agreed		Observed agreement	Kappa	95% CI Kappa
			Yes	No			
Supraspinatus-tendinopathy	0.15	0.11–0.21	37	295	0.89	0.59	0.47–0.70
Infraspinatus-tendinopathy	0.08	0.03–0.10	16	331	0.93	0.53	0.34–0.68
Subscapularis-tendinopathy	0.07	0.06–0.08	18	337	0.96	0.68	0.53–0.82
Subacromial bursitis chronic	0.15	0.06–0.19	31	291	0.86	0.46	0.33–0.58
Glenohumeral capsulitis	0.35	0.31–0.44	100	214	0.84	0.66	0.57–0.73
Acromio-clavicular joint lesion	0.06	0.05–0.10	19	343	0.97	0.78	0.61–0.90
Neck or shoulder girdle lesion	0.28	0.26–0.29	93	255	0.94	0.84	0.78–0.90
Rotator cuff tendinopathy <sup>a</sup>	0.23	0.18–0.29	63	264	0.88	0.66	0.57–0.75
Bursitis <sup>b</sup>	0.17	0.15–0.19	33	279	0.84	0.43	0.30–0.55

Diagnostic agreement is presented as a number of the 372 pairs that agreed on a diagnosis or not, and observed proportion of agreement. Chance-corrected agreement (kappa) are presented with 95% confidence interval (CI).

<sup>a</sup>Collapsed variable including supraspinatus-, infraspinatus-, and subscapularis-tendinopathies.

<sup>b</sup>Collapsed variable including includes chronic and acute subacromial and subdeltoid bursitis.

acromioclavicular joint lesion diagnoses, and subscapularis-tendinopathy, moderate agreement for supra- and infraspinatus-tendinopathy. Agreements on diagnoses were generally better than on individual tests.

### **Strengths and weaknesses**

One of the strengths of this study is that it is a reflection of primary health care practice, as called for in the *JAMA* review by Hermans et al. in 2013.[21] “Ordinary” patients, with heavy symptom load and long duration of complaints, were recruited in a sufficient number to reach enough statistical power for the kappa statistics. The four clinicians registered their results consecutively in a standardized clinical recording form, without any possibility for changing view after discussions with peers. Without pre-calibration, this study may not reflect the highest possible scores, but probably reflects achievable reliability using the diagnostic principles of orthopaedic medicine and STTs in primary health care.

There are, however, limitations to our study. The examiners were limited in using accessory test and in providing diagnosis beyond the items provided in the questionnaire. Recall error upon specific tests may be possible as data input provided by the clinicians were at the end of each examination. A focus on individual tests with consecutive registration instead of the inherent picture might have changed the results for specific tests. However, the results are in line with results for individual tests for impingement,[11] mobility disorders,[22] and other shoulder conditions [11] that all demonstrated low reliability and caution in diagnosing based on single tests. Repetitive examinations in a tight time schedule might provoke symptoms and thereby reduce agreement. The workload by examining over 60 patients in two days was substantial, but we could not demonstrate reduced agreements at the end of the days.

### **Findings in relation to other studies**

The results indicate that the tests included in the basic functional examination combined with a detailed medical history, constitute a useful systematic approach to clinical shoulder examination, and the results correlate well with other studies investigating this diagnostic method.[14,16] As opposed to Hanchard et al.,[14] and Pellicchia et al.[16] no pre-calibration training was conducted prior to our study to reflect the “real life situation”. Furthermore, the participating clinicians work in four different workplaces in two different countries,

which altogether make it more challenging to obtain good agreement between assessors.

In this study, we defined three diagnostic levels: the individual tests, the individual diagnoses, and the diagnostic groups. The highest kappa scores were obtained for the diagnostic groups, and the lowest for the individual tests. The discrepancy in kappa scores between diagnostic levels is an important finding. It demonstrates that when making a diagnosis, experienced clinicians use comprehensive clinical reasoning skills in addition to information from the summary of specific tests.[23,24] This was illustrated in a qualitative study where diagnostic reasoning involved both pattern recognition and hypothetico-deductive reasoning on assessment of patients with shoulder pain.[25] Myer et al. recently highlighted in their user’s guide to examination of the shoulder that the individual tests are best used in the context of a comprehensive history and physical examination, [26] a notion which was supported by Hegedus et al. in their meta-analysis from 2012.[12]

A lower kappa score of 0.46 was obtained in this study for chronic subacromial bursitis. The study by Hanchard et al. obtained similar kappa scores for bursitis, ranging between 0.35 and 0.58.[14] The lower kappa scores might be explained by the mixed clinical picture of pain on some passive movements and pain on some resisted movements, with or without painful arc in bursitis.[27] Cyriax considered this pattern to be an “incomprehensible bursitis”. The main difficulty with chronic subdeltoid bursitis remains the heterogeneity of the clinical pattern.

Hanchard et al. [10] considered rotator cuff tendinopathies as one group, and obtained interobserver agreements ranging between 0.71 and 0.79 compared to the observed agreement of 0.88, and the kappa score of 0.66 for rotator cuff tendinopathy as one group in this study. The complexity of diagnosis of rotator cuff tendinopathies, and their relationship with bursal pathology has been highlighted,[28] indicating the challenges in relation to precise diagnosis between these two conditions. In addition, repetitive testing of shoulder may have influenced test results by provoking pain through the test procedures.

Glenohumeral capsulitis is characterized by painful, gradual loss of active and passive shoulder motions. The capsular pattern at the shoulder joint is described as proportional limitation of the three passive scapulohumeral movements; there is some limitation of abduction, more limitation of external rotation, and less limitation of internal rotation.[27] As the literature suggests that other shoulder pathology can produce a similar clinical picture, a precise diagnosis may also here be

a challenge.[29] Still, a kappa score of 0.66 in this study and kappa scores in the range between 0.63 and 0.82 in the study by Hanchard et al. [14] indicate that trained clinicians may identify this diagnosis with substantial accuracy in primary health care. The good agreement in testing passive abduction and lateral rotation probably contributes to the diagnostic agreement.

The true prevalence of acromioclavicular joint lesion is unknown [30] and we are not aware of other studies reporting intertester reliability on this lesion. The kappa score of 0.78 obtained for the acromioclavicular joint lesion is substantial, although also here diagnosis is considered challenging.[30]

### **Reliability and validity**

Our results revealed moderate to good intertester agreement on rotator cuff tendinopathies, indicating that validity might be a challenge. Calculation of agreement based on figures presented in the Cochrane review [10] (Figures 5, 9, and 10), indicate agreement, with relatively low kappa values (low to fair/good),[19] between different tests for rotator cuff tendinopathy and arthroscopic reference standard. The test agreement in the Cochrane review [10] is in line with the kappa values found for individual STTs (0.15–0.45) in our study, except for our results for limitation of passive lateral rotation and passive abduction which performed better (0.66 and 0.70, respectively). A recent Norwegian study indicates that standardization and use of plurimeters might improve reliability of the limitation test.[31]

### **Implications for practice**

Although specific diagnosis of shoulder lesions in general practice remains challenging, this study supports that the diagnostic principles of orthopaedic medicine and STTs can be a valuable and manageable diagnostic tool for GPs and physical therapists in primary health care settings. Better agreement on diagnoses than on specific tests underscores the necessity of systematic functional examination and history to recognize clinical pictures. Still, the reliability shows variation in diagnoses that challenge specific treatments and comparisons of results. Further research is necessary on how to improve the diagnostic system for classification of shoulder complaints.

### **Acknowledgements**

We are thankful for the clinicians who referred the shoulder patients to our study, to the patients who willingly were

examined four times, and to the clinicians (Andersson Thomas, Gustavsson Tomas, Ingemann Are, and Røsand Tom) for sharing their clinical skills with us during two intensive days.

### **Ethical approval**

The study was approved by the Data Protection Official for Research. The regional ethics committee considered the study to be outside their assignment.

### **Disclosure statement**

The authors declare that they have no competing interests.

### **Funding information**

The project was funded by grants from UiT the Arctic University of Norway.

### **References**

- [1] Hagen K, Linde M, Heuch I, et al. Increasing prevalence of chronic musculoskeletal complaints. A large 11-year follow-up in the general population (HUNT 2 and 3). *Pain Med.* 2011;12:1657–1666.
- [2] Roe Y, Soberg HL, Bautz-Holter E, et al. A systematic review of measures of shoulder pain and functioning using the International classification of functioning, disability and health (ICF). *BMC Musculoskelet Disord.* 2013;14:73.
- [3] Virta L, Joranger P, Brox JI, et al. Costs of shoulder pain and resource use in primary health care: a cost-of-illness study in Sweden. *BMC Musculoskelet Disord.* 2012;13:17
- [4] van der Heijden GJ, van der Windt DA, de Winter AF. Physiotherapy for patients with soft tissue shoulder disorders: a systematic review of randomised clinical trials. *BMJ.* 1997;315:25–30.
- [5] Monrad SU, Zeller JL, Craig CL, et al. Musculoskeletal education in US medical schools: lessons from the past and suggestions for the future. *Curr Rev Musculoskelet Med.* 2011;4:91–98.
- [6] Kooijman M, Swinkels I, van Dijk C, et al. Patients with shoulder syndromes in general and physiotherapy practice: an observational study. *BMC Musculoskelet Disord.* 2013;14:128.
- [7] Varela E, Valero R, Kucukdeveci AA, et al. Committee U-PSPP. Shoulder pain management. The role of physical and rehabilitation medicine physicians. The European perspective based on the best evidence. A paper by the UEMS-PRM Section Professional Practice Committee. *Eur J Phys Rehabil Med.* 2013;49:743–751.
- [8] Loeffler BJ, Brown SL, D'Alessandro DF, et al. Incidence of false positive rotator cuff pathology in MRIs of patients with adhesive capsulitis. *Orthopedics.* 2011;34:362.
- [9] Moen MH, de Vos RJ, Ellenbecker TS, et al. Clinical tests in shoulder examination: how to perform them. *Br J Sports Med.* 2010;44:370–375.

- [10] Hanchard NC, Lenza M, Handoll HH, et al. Physical tests for shoulder impingements and local lesions of bursa, tendon or labrum that may accompany impingement. *Cochrane Database Syst Rev*. 2013;4:CD007427.
- [11] Hegedus EJ, Goode A, Campbell S, et al. Physical examination tests of the shoulder: a systematic review with meta-analysis of individual tests. *Br J Sports Med*. 2008;42:80–92. discussion 92.
- [12] Hegedus EJ, Goode AP, Cook CE, et al. Which physical examination tests provide clinicians with the most value when examining the shoulder? Update of a systematic review with meta-analysis of individual tests. *Br J Sports Med*. 2012;46:964–978.
- [13] Carter T, Hall H, McIntosh G, et al. Intertester reliability of a classification system for shoulder pain. *Physiotherapy*. 2012;98:40–46.
- [14] Hanchard NC, Howe TE, Gilbert MM. Diagnosis of shoulder pain by history and selective tissue tension: agreement between assessors. *J Orthop Sports Phys Ther*. 2005;35:147–153.
- [15] Indelicato PA. Presidential address of the American Orthopaedic Society for Sports Medicine: how good do you want to be . . . aim high. *Am J Sports Med*. 2012;40:2690–2693.
- [16] Pellecchia GL, Paolino J, Connell J. Intertester reliability of the cyriax evaluation in assessing patients with shoulder pain. *J Orthop Sports Phys Ther*. 1996;23:34–38.
- [17] Ombregt L. *A System of orthopaedic medicine*. 3rd ed. Edinburgh: Churchill Livingstone; 2013.
- [18] Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther*. 2005;85:257–268.
- [19] Altman DG. *Practical statistics for medical research*. 1st ed. London: Chapman and Hall; 1991.
- [20] Roberts C, McNamee R. Assessing the reliability of ordered categorical scales using kappa-type statistics. *Stat Methods Med Res*. 2005;14:493–514.
- [21] Hermans J, Luime JJ, Meuffels DE, et al. Does this patient with shoulder pain have rotator cuff disease?: The Rational Clinical Examination systematic review. *JAMA*. 2013;310:837–847.
- [22] Kolber MJ, Fuller C, Marshall J, et al. The reliability and concurrent validity of scapular plane shoulder elevation measurements using a digital inclinometer and goniometer. *Physiother Theory Pract*. 2012;28:161–168.
- [23] Levine D, Bleakley A. Rethinking clinical reasoning. *Med Educ*. 2013;47:745–746.
- [24] Sniderman AD, LaChapelle KJ, Rachon NA, et al. The necessity for clinical reasoning in the era of evidence-based medicine. *Mayo Clin Proc*. 2013;88:1108–1114.
- [25] May S, Greasley A, Reeve S, et al. Expert therapists use specific clinical reasoning processes in the assessment and management of patients with shoulder pain: a qualitative study. *Aust J Physiother*. 2008;54:261–266.
- [26] Myer CA, Hegedus EJ, Tarara DT, et al. A user's guide to performance of the best shoulder physical examination tests. *Br J Sports Med*. 2013;47:903–907.
- [27] Ombregt LBP, Veer HJT. *A system of orthopaedic medicine/with accompanying CD-ROM*. 2nd ed. London: Churchill Livingstone. 2003.
- [28] Lewis JS. Rotator cuff tendinopathy: a model for the continuum of pathology and related management. *Br J Sports Med*. 2010;44:918–923.
- [29] Neviasser AS, Neviasser RJ. Adhesive capsulitis of the shoulder. *J Am Acad Orthop Surg*. 2011;19:536–542.
- [30] Andreani L, Bonicoli E, Parchi P, et al. Acromio-clavicular repair using two different techniques. *Eur J Orthop Surg Traumatol*. 2014;24:237–242.
- [31] Sharma SP, Baerheim A, Kvale A. Passive range of motion in patients with adhesive shoulder capsulitis, an intertester reliability study over eight weeks. *BMC Musculoskelet Disord*. 2015;16:37.