

Intertester reliability of a classification system for shoulder pain

Tom Carter^a, Hamilton Hall^{a,b}, Greg McIntosh^{a,*}, John Murphy^c,
Janice MacDougall^d, Christina Boyle^a

^a CBI Health Group, Toronto, Ontario, Canada

^b Department of Surgery, University of Toronto, Toronto, Ontario, Canada

^c Saskatoon Health Region, Saskatoon, Saskatchewan, Canada

^d CBI Health Group, Saskatoon, Saskatchewan, Canada

Abstract

Objective To describe and determine the intertester reliability of a newly developed classification system of shoulder syndrome recognition.

Design Intertester reliability study.

Setting Fourteen primary care physiotherapy clinics.

Participants Two hundred and fifty-five patients with shoulder pain. Inclusion criterion: presence of shoulder pain arising within the glenohumeral or associated joints and structures. Exclusion criteria: previous shoulder surgery, surgical candidates, recognised malignancy, systemic illness, or concurrent cervical pain and/or radiculopathy.

Intervention Examiners were 55 physiotherapists who were arranged in pairs; each patient received two independent and blinded assessments, one by each of the paired physiotherapists. This shoulder classification approach contains three main clinical syndromes: Pattern 1 (impingement pain), Pattern 2 (acromioclavicular joint pain) and Pattern 3 (shoulder pain: frozen shoulder, glenohumeral arthritis, massive cuff tear, subscapularis tear, painful laxity, post-traumatic instability, internal derangement).

Main outcome measures Percentage agreement and Cohen's kappa coefficient.

Results The mean age of patients was 46.6 years (standard deviation 16.3, range 16 to 86), and 57% were male. Physiotherapists agreed on the pattern of shoulder pain for 205 of the 255 shoulders assessed (agreement rate 80%); the kappa coefficient was 0.664 (95% confidence interval 0.622 to 0.706; $P < 0.001$). Of the 205 agreements, Pattern 1 was the most common condition; physiotherapists agreed on this pattern for 139 patients (68%). Both physiotherapists diagnosed Pattern 2 for 20 patients and Pattern 3 for 46 patients.

Conclusion This clearly defined system uses key elements of the history and examination to classify patients with shoulder pain. The kappa coefficient denotes good reproducibility.

© 2011 Chartered Society of Physiotherapy. Published by Elsevier Ltd. All rights reserved.

Keywords: Shoulder patterns; Kappa; Reliability; Clinical syndromes

Introduction

Shoulder disorders constitute a medical and economic challenge to society [1–5]. Determining the diagnosis based on anatomical location of the pain generator is not a straightforward process [6] because the pathophysiology of shoulder pain is unclear [5], not all structural failures produce symptoms [7–9], and there is a weak correlation between medical imaging and symptoms [8,10–16]. However, some method

of classification is essential because, in theory, diagnostic labels facilitate communication, enhance treatment success and improve outcomes [5,17]. Appropriate patient classification is a necessary preliminary step in defining and managing clinical problems [18].

Only a few researchers have examined shoulder assessment and classification methods for reliability [8,19]; generally, the intertester agreement rate has been poor, with kappa coefficients ranging from 0.31 to 0.45 [6,20–22]. Only Pellecchia *et al.* [20] achieved impressive results (91% agreement; kappa = 0.88) using the Cyriax method of shoulder pain classification. Cyriax's schema is a patho-anatomical classification with nine possible categories. The system involves a

* Corresponding author at: 153 Oliver Place, Oakville, Ontario, Canada L6H 1K9. Tel.: +1 905 844 3871; fax: +1 416 231 0091.

E-mail address: gmcintosh@cbi.ca (G. McIntosh).

detailed patient history, a preliminary examination of active movements to eliminate the neck and elbow as sources of pain, and a detailed examination of the affected shoulder to diagnose soft tissue lesions. Pellecchia *et al.* [20] described specific details of the Cyriax system. Unfortunately, this study of the Cyriax method had a very small sample size ($n = 21$) and only used two therapists to assess agreement. The lack of any patient demographic information limits the generalisability of this study. Furthermore, a subsequent study did not show good reliability with the Cyriax method [6].

Smidt and Green [23] concluded that the reproducibility of shoulder classification is inadequate. Progress has been hampered by failure to reach consensus on appropriate diagnostic criteria and the lack of a uniform system of categorising shoulder patients into clinically meaningful subgroups. In addition, consistent descriptions, uniform definitions and specific physical findings are not used [17]; one researcher concluded that there is ‘terminological chaos in the field of shoulder pain’ [24].

Recognising these difficulties prompted this research project to determine the reliability of a simple yet comprehensive approach to organising patterns of shoulder pain. The purpose of this study was to determine the intertester reliability of a newly developed shoulder classification system. The research question was: ‘using this examination-based classification system, can physiotherapists assessing the same shoulder condition independently identify the same syndrome?’

Methods

This was a reliability study rather than a true diagnostic study; thus, the STARD guidelines for diagnostic studies do not strictly apply to all aspects of this project. However, since there are no set criteria to judge the quality of reliability studies [5], this study followed the reliability checklist proposed by May *et al.* [25] and, where appropriate, the STARD guidelines [26].

Sample

This was a multiclinic study of 255 patients, conducted at 12 physiotherapy clinics across three provinces (Ontario, Saskatchewan, Alberta) in Canada and at two clinics in New Zealand. The study began on 6 February 2007 and ended on 18 December 2008. Patient confidentiality and privacy were ensured for all aspects of the study design and data analysis. All subjects provided written informed consent and patients’ rights were protected.

Part of evaluating a classification system is assessing its comprehensiveness, and it is advantageous to have few inclusion/exclusion criteria to maximise external validity [27]. The inclusion criterion was the presence of shoulder pain arising within the glenohumeral or associated joints and structures. Individuals with previous shoulder surgery, surgical candidates, recognised malignancy, systemic illness, or identified concurrent cervical pain and/or radiculopathy were excluded.

Examiners and classification

Based on their frustration with the lack of diagnostic consensus in the literature, two of the authors (JM, JMc) originally developed a shoulder classification system in 2004. After a 3-month trial period, they consulted with the third author (HH) to make improvements. In 2005, after another 6-month trial period, the system was presented to 12 senior physiotherapists for feedback; this meeting resulted in further refinements to the system. Those who attended the meeting embarked on a 1-year trial to assess the clinical utility of the system. To ensure face, content and construct validity in the development process, strict operational definitions were documented and the classification system was revised every 3 months during the year based on clinicians’ feedback from approximately 100 patients. In late 2006, two orthopaedists modified the system further and shaped it into the current form.

The examiners were 55 physiotherapists trained to use the new method of shoulder classification. Prior to the study, each clinic received an instruction package. The contents consisted of: (1) directions for staff in charge of scheduling appointments; (2) an information page for patients; (3) patient consent and release forms; (4) ballot forms for participating physiotherapists to indicate their conclusions; and (5) a one-page synopsis of the study methods.

The essential categories in this classification system are determined by pain location, patient age, mechanism of injury, and aggravating factors determined via specific provocative tests. The literature indicates that the validity and reliability of shoulder testing based primarily on physical examination are poor [5,17]. For this reason, the current classification system gives the history precedence over the physical examination. Since the patient’s history (pain location, age and mechanism of injury) is the essential component of this approach, the physiotherapists used a pragmatic approach in the physical examination; they performed any provocative tests they deemed appropriate based on the history.

This shoulder classification system only defines discrete clinical syndromes and makes no attempt to describe the shoulder pathology definitively. It contains three main categories or patterns of shoulder pain. The label for each pattern signifies a presumed pain source, which is to be addressed in the initial management. Table 1 summarises the patterns of shoulder pain relative to pain location, patient age, mechanism of injury and physical examination findings. The clinical syndromes, presented as patterns of pain, are designated as:

- Pattern 1: impingement pain (pain felt predominantly over the antero-lateral deltoid area).
- Pattern 2: acromioclavicular joint pain (pain felt directly over the acromioclavicular joint).
- Pattern 3: shoulder pain (pain may be felt over the anterior or posterior aspect of the joint, or may occur simultaneously in both locations).

Table 1
Clinical syndromes of shoulder pain based on history and physical examination.

Pattern of shoulder pain	Pain location	Mechanism	Approximate age (years)	Main physical examination findings
Pattern 1	Deltoid	Traumatic or non-traumatic	35+	Tenderness over greater tuberosity. Painful arc. Pain inhibition with apparent weakness. Able to actively abduct arm from side.
Pattern 2	Acromioclavicular joint	Traumatic Non-traumatic	Any age 50+	Joint prominence. Tender on palpation of acromioclavicular joint. Pain with resisted horizontal extension.
Pattern 3 Frozen shoulder	Through the joint area	Traumatic or non-traumatic	40 to 60	Painful with all movements. Loss of ROM (loss of ER > abduction/IR). Progresses to global loss of ROM. Active and passive ROM equally limited.
Glenohumeral osteoarthritis	Through the joint area	Gradual	50+	Decreased active ROM (usually flexion). Global loss of movement. Painful crepitus.
Complete cuff tear	Through the joint area	Recognised event	60+	Normal passive ROM. Unable to initiate abduction with arm at side but possible with arm above 90°.
Subscapularis tear	Through the joint area	Traumatic, event-forced external rotation	40+	Tenderness over anterior shoulder. Reduced ROM. Weakness of internal rotation.
Painful laxity	Through the joint area	Minimal but recognised trauma	Teens to 20s	Hypermobility shoulder. May have generalised laxity.
Post-traumatic instability	Through the joint area	Following traumatic dislocation	Teens to 20s	ROM usually normal. Positive apprehension sign.
Internal derangement	Through the joint area	Repeated throwing motion	18 to 30	No findings of impingement.

Pattern 1, impingement pain; Pattern 2, acromioclavicular joint pain; Pattern 3, shoulder pain; ER, external rotation; IR, internal rotation; ROM, range of motion.

For convenience, the pain location is described as ‘through the joint area’ in Table 1.

Pattern 3 is subdivided into seven painful conditions:

- frozen shoulder;
- glenohumeral osteoarthritis;
- complete cuff tear;
- subscapularis tear;
- painful laxity;
- post-traumatic instability; and
- internal derangement.

Pattern 3 represents a range of potential pain generators. The individual conditions within Pattern 3 are considerably less common than those for Patterns 1 and 2, but as each is unique, they are listed separately. Based on the fact that they all share a similar, common pain location, they are categorised together in one pattern (Pattern 3). In this reliability study, it was not sufficient for physiotherapists to simply state that a patient was Pattern 3; they were required to identify one of the seven conditions within the pattern.

Procedure

Physiotherapists were arranged in pairs. Following consent to participate in the study, an eligible patient was scheduled for two assessments, one by each of the paired physiotherapists. Each physiotherapist independently (and blinded to each other’s assessment) completed a standardised shoulder examination form to record the pertinent information. Each physiotherapist also completed a simple ballot to indicate the pattern of pain; this determination was made immediately following the assessment without consultation with the other physiotherapist. The completed ballots were sealed in individual envelopes and forwarded to the research co-ordinator.

To maintain internal validity and because an intention of this study was to examine the agreement rate of this classification in clinical practice, a minimal departure from usual daily clinic operation was an essential requirement. For logistical and scheduling reasons, the two assessments were either performed in tandem or as completely separate assessments. In

75% of the cases, patients received both assessments on the same day. For the tandem assessments, both examiners were in the room with the patient for the history taking. Physiotherapist 1 provided the questions and interaction with the patient while Physiotherapist 2 listened and observed. Physiotherapist 2 refrained from interfering with the patient's history. During this phase of the assessment, there was no conversation between the two physiotherapists and no contact between Physiotherapist 2 and the patient. The designations of Physiotherapists 1 and 2 were assigned at random. Following the history, Physiotherapist 2 left the room. Physiotherapist 1 then completed a physical examination consisting of a cervical screen, observation of the shoulder, assessment of active and passive range of shoulder motion, and provocative testing. After Physiotherapist 1 had completed the physical examination, the patient had a second physical examination by Physiotherapist 2, not attended by Physiotherapist 1. Patients who had completely separate (not tandem) assessments received two separate physical examinations, but in addition were asked to provide two separate histories.

As the examinations were performed within functioning clinics, scheduling constraints required that 25% of the patients received their separate assessments on different days; the average time between assessments was less than 1 day for 75% (174/233) of the participants (range 0 to 23 days).

Following both assessments, either Physiotherapist 1 or 2 provided ongoing treatment depending on schedule availability.

Data analysis

Cohen's kappa coefficient and percentage agreement were calculated; kappa should be presented in tandem with other statistics to facilitate better interpretation of any classification system [28].

Three assumptions must be met to use the kappa statistic properly [29,30]:

- (1) the data should be categorical;
- (2) examiners must make their decisions independently (examiners should not have influence over each other); and
- (3) the patient's information should be independent.

Assumption 3 is met when all patients being rated represent independent data points; the assumption is violated when knowing a patient's initial response to some part of an assessment provides information about subsequent responses (i.e. the patient's data points correlate in some way).

The current study was constructed carefully to meet these statistical assumptions. Specifically:

- (1) the patterns of pain system is categorical, not ordinal or scale;
- (2) each physiotherapist conducted his/her physical examinations independently in a separate room without the second physiotherapist present. Subsequently, without

consultation with the other examiner, each physiotherapist completed a ballot indicating the pattern of shoulder pain; one physiotherapist was not simply confirming or refuting the opinion of another; and

- (3) patient information was only obtained at the initial assessment; a pre–post study design was not utilised.

Analyses were conducted using Statistical Package for the Social Sciences Version 15.0.1 (IBM Corporation, NY, USA).

Sample size

The design of a study determines the appropriate analysis to perform sample size calculations. Sim and Wright's [30] table for sample size estimation was used to determine the number of subjects required. The sample size calculation estimated that 255 patients were required in order to provide sufficient power (0.80) for statistical analysis of the findings, based on the following criteria:

- (1) a two-tailed test at a level of significance of $\alpha = 0.05$;
- (2) minimal kappa of clinical significance of 0.40;
- (3) expected kappa between clinicians of 0.70; and
- (4) expected agreement between clinicians of 0.70.

Results

The study consisted of 275 patients; eight only had one assessment and 12 had pain unrelated to the shoulder, resulting in a final sample size of 255. The average patient age was 46.6 years (standard deviation 16.3, range 16 to 86), and males represented 57% of the sample.

Physiotherapists agreed on the pattern of shoulder pain for 205 of the 255 shoulders assessed (agreement rate 80%, kappa coefficient 0.664, 95% confidence interval 0.622 to 0.706; $P < 0.001$). Of the 205 agreements, Pattern 1 was by far the most common condition; physiotherapists agreed that Pattern 1 was the pain pattern for 139 patients, Pattern 2 for 20 patients and Pattern 3 for 46 patients. Specific analysis within Pattern 3 revealed that frozen shoulder ($n = 15$) and post-traumatic instability ($n = 15$) were most common; the remaining conditions were rare in this clinical setting.

Table 2 provides a matrix of where the 205 agreements (in bold) and 50 disagreements lay. Physiotherapist 1's choices are read vertically and Physiotherapist 2's choices are read horizontally. An example of how to read the table is provided at the bottom of Table 2.

Discussion

A fundamental requirement of an effective classification system is good intertester reliability. Studies of intertester reliability are more accurately described as investigations of agreement between raters [29]. The current study achieved an 80% agreement rate. There were disagreements. Confus-

Table 2
Matrix of the 205 agreements (in bold) and 50 disagreements by pattern of shoulder pain*.

Pattern of shoulder pain	Physiotherapist 1										Total
	Pattern 1	Pattern 2	Pattern 3								
			FS	GH	CCT	SST	PL	PTI	ID	NSP	
Physiotherapist 2											
Pattern 1	139	4	2	1	4	3	2			5	160
Pattern 2	2	20								1	23
Pattern 3											
FS	7	1	15								23
GH			1	1							2
CCT	2				7						9
SST	1					3		1			5
PL	6						4	2	1		13
PTI	1							15			16
ID		1							1		2
NSP	2										2
Total	160	26	18	2	11	6	6	18	2	6	255

Pattern 1, impingement pain; Pattern 2, acromioclavicular joint pain; Pattern 3, shoulder pain; FS, frozen shoulder; GH, glenohumeral arthritis; CCT, complete cuff tear; SST, subscapularis tear; PL, painful laxity; PTI, post-traumatic instability; ID, internal derangement; NSP, non-shoulder pain.

* Physiotherapist 1's choices are read vertically and Physiotherapist 2's choices are read horizontally. For example, Pattern 2 was identified 26 times by Physiotherapist 1 and 23 times by Physiotherapist 2. In 20 cases, there was agreement (in bold); in four cases, Physiotherapist 1 diagnosed Pattern 2 while Physiotherapist 2 diagnosed Pattern 1. In another case, Physiotherapist 2 diagnosed FS. In two cases, Physiotherapist 2 diagnosed Pattern 2 but Physiotherapist 1 diagnosed Pattern 1. In one case, Physiotherapist 1 believed the pain was not related to the shoulder.

ing frozen shoulder with Pattern 1 or Pattern 2 reflects a physiotherapist's inability to accurately identify a patient's site of dominant pain. Although this classification is simple, the application is not necessarily easy because the system requires careful attention to the details within the history and physical examination.

More importantly, the kappa coefficient was 0.66. Unfortunately, there is no clearly accepted interpretation of the kappa statistic. Several researchers [29,31] suggest that a kappa coefficient of 0.66 represents good agreement. Others [32,33] view a kappa coefficient of 0.66 as substantial agreement. While there is debate on what actually constitutes good agreement, the shoulder pain classification system used in this study achieved a minimum of good reliability.

When the prevalence of one condition within a classification system is high, as opposed to subjects spread more evenly across categories, the consequence is a more homogeneous sample; lower kappa values usually result [29]. This classification study had a predominance of Pattern 1 patients (68%); the other patterns were somewhat rare in comparison. Although the high percentage of Pattern 1 patients may have led to some homogeneity, the kappa coefficient was higher than that found in other studies [16,21,22], which emphasises the discrimination possible with this system.

Musculoskeletal assessment of the shoulder is usually based on a premise that it is possible to isolate individual structures and apply a mechanical procedure that compresses, stretches, contracts or in some way irritates the tissue of interest. However, involvement of adjacent structures is likely, and perceived weakness may be better explained as pain inhibition rather than the result of structural failure [34]. The reliability of a physical examination to identify specific patho-anatomical mechanisms resulting in shoulder symp-

oms is poor [6,16], and the diagnostic validity is only moderate [35].

The commonly used special shoulder symptom provocation tests do not contribute to a precise structural diagnosis or isolate specific structures around the shoulder [8]. For this reason, the current classification system emphasises that the history takes precedence over the physical examination. This system of shoulder pain pattern identification emphasises a small number of reliable, clearly defined, key components of the history that allow syndrome categorisation. The role of the physical examination is to support or refute a determination based on the patient's story.

A limitation of this study includes the potential for bias. Since the treatment to be provided by the clinics involved in this study is an active rehabilitation programme, physicians may decide that for profoundly disabled or minimally restricted patients, this form of management is not appropriate, thereby creating a referral filter bias. However, evidence from a pilot project involving other rehabilitation providers [36] suggests that patient pain, severity and functional status measures for this clinic system are similar to six other non-related physiotherapy facilities in Ontario. Therefore, it is likely that there was little, if any, selection bias in this study.

Logistically, it was not possible for all patients to receive both assessments in succession. Approximately 75% of the sample had both assessments on the same day; for the others, the average time between assessments was less than 1 day. This aspect of the study design reduced the potential for bias.

The issue of criterion validity cannot be properly addressed because of the lack of a diagnostic gold standard. Traditionally, the gold standards have been magnetic resonance imaging, ultrasound or surgical exploration, but weak correlation between symptoms and structures inhibits

diagnostic validity. Furthermore, a lack of consensus on appropriate diagnostic criteria adds to the complexity of shoulder pain assessment [16]. These diagnostic limitations highlight the need for a model that is not based on pathology but rather on clinical presentation. In fact, May *et al.* [5] recommend that clinicians should abandon the diagnostic pathological model. This classification of syndromes rather than specific aetiologies subscribes to that recommendation.

Strengths of this study include standardisation of the shoulder assessment methods. By using fully integrated clinics with the same centrally co-ordinated data collection tools and philosophy of assessment, this study reduced the potential for unexpected bias or inadequate data accumulation. This classification system suggests that if common patterns of shoulder pain are clinically recognised, they could form the basis for triaging and primary non-operative management. It should be stressed that this history-based classification is only intended to direct the initial intervention and makes no attempt to delineate subsequent treatment options. Further exhaustive, costly and perhaps unnecessary diagnostic evaluations might, however, be deferred or employed only after a failure to respond to primary care. The early use of such investigations could also be reserved for clinical presentations that do not fit one of the three described patterns of shoulder pain.

Conclusion

This clearly defined system primarily uses key elements of the history with a confirmatory physical examination to classify patients with shoulder pain. This method demonstrated a high degree of intertester agreement, correctly categorising 80% of patients with shoulder pain. The kappa coefficient of 0.664 denotes good reproducibility.

This system offers a simplified approach to the primary assessment of shoulder pain. It avoids prematurely confronting the complex pathophysiological mechanisms that may contribute to the pain, while providing a description of the problem sufficient to guide initial management.

Acknowledgements

The authors thank all the physiotherapists who participated in Edmonton (Alberta), Saskatoon (Saskatchewan), London, Woodbridge, Sarnia, Mississauga, Etobicoke, Ottawa (Ontario) and New Zealand. In particular, the authors thank Lloyd Vinish, Chris Gregg and Jeff Vanderslagt for their extended contributions.

Ethical approval: Research Ethics Board of Sunnybrook Health Sciences Centre (Project Id. No.: 430-2006, approval date: 11/09/2006).

Funding: CBI Health Group.

Conflict of interest: None declared.

References

- [1] da Costa BR, Armijo-Olivo S, Gadotti I, Warren S, Reid DC, Magee DJ. Reliability of scapular positioning measurement procedure using the Palpation Meter (PALM). *Physiotherapy* 2010;96:59–67.
- [2] Geraets JJ, Goossens ME, De Bruijn CP, De Groot IJ, Koke AJ, Pelt RA, *et al.* Cost-effectiveness of a graded exercise therapy program for patients with chronic shoulder complaints. *Int J Technol Assess Health Care* 2006;22:76–83.
- [3] Luime JJ, Koes BW, Hendriksen IJ, Burdorf A, Verhagen AP, Miedema HS, *et al.* Prevalence and incidence of shoulder pain in the general population: a systematic review. *Scand J Rheumatol* 2004;33:73–81.
- [4] Kuijpers T, Van Tulder MW, Van Der Heijden GJ, Bouter LM, Van Der Windt DA. Costs of shoulder pain in primary care consulters: a prospective cohort study in The Netherlands. *BMC Musculoskelet Dis* 2006;7:83–90.
- [5] May S, Chance-Larsen K, Littlewood C, Lomas D, Saad M. Reliability of physical examination tests used in the assessment of patients with shoulder problems: a systematic review. *Physiotherapy* 2010;96:179–90.
- [6] Connor PM, Banks DM, Tyson AB, Coumas JS, D'Alessandro DF. Magnetic resonance imaging of the asymptomatic shoulder of overhead athletes: a 5-year follow-up study. *Am J Sports Med* 2003;31:724–7.
- [7] Lewis JS. Rotator cuff tendinopathy/subacromial impingement syndrome: is it time for a new method of assessment? *Br J Sports Med* 2009;43:259–64.
- [8] Miniaci A, Mascia AT, Salonen DC, Becker EJ. Magnetic resonance imaging of the shoulder in asymptomatic professional baseball pitchers. *Am J Sports Med* 2002;30:66–73.
- [9] Fredericson M, Ho C, Waite B, Jennings F, Peterson J, Williams C, *et al.* Magnetic resonance imaging abnormalities in the shoulder and wrist joints of asymptomatic elite athletes. *Phys Med Rehabil* 2009;1:107–16.
- [10] Reuter RM, Hiller WD, Ainge GR, Brown DW, Dierenfield L, Shellock FG, *et al.* Ironman triathletes: MRI assessment of the shoulder. *Skeletal Radiol* 2008;37:737–41.
- [11] Krief OP, Huguet D. Shoulder pain and disability: comparison with MR findings. *AJR Am J Roentgenol* 2006;186:1234–9.
- [12] Bradley MP, Tung G, Green A. Overutilization of shoulder magnetic resonance imaging as a diagnostic screening tool in patients with chronic shoulder pain. *J Shoulder Elbow Surg* 2005;14:233–7.
- [13] Hegedus EJ, Goode A, Campbell S, Morin A, Tamaddoni M, Moorman CT, *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.
- [14] Miller CA, Forrester GA, Lewis JS. The validity of the lag signs in diagnosing full-thickness tears of the rotator cuff: a preliminary investigation. *Arch Phys Med Rehabil* 2008;89:1162–8.
- [15] Park HB, Yokota A, Gill HS, El Rassi G, McFarland EG. Diagnostic accuracy of clinical tests for the different degrees of subacromial impingement syndrome. *J Bone Joint Surg Am* 2005;87:1446–55.
- [16] de Winter AF, Jans MP, Scholten RJ, Deville W, van Schaardenburg D, Bouter LM. Diagnostic classification of shoulder disorders: interobserver agreement and determinants of disagreement. *Ann Rheum Dis* 1999;58:272–7.
- [17] Schellinghouth JM, Verhagen AP, Thomas S, Koes BW. Lack of uniformity in diagnostic labeling of shoulder pain: time for a different approach. *Man Ther* 2008;13:478–83.
- [18] Nachemson A, Andersson GBJ. Classification of low-back pain. *Scand J Work Environ Health* 1982;8:134–6.
- [19] Lewis JS, Wright C, Green A. Subacromial impingement syndrome: the effect of changing posture on shoulder range of movement. *J Orthop Sports Phys Ther* 2005;35:72–87.
- [20] Pellecchia GL, Paolino J, Connell J. Intertester reliability of the Cyriax evaluation in assessing patients with shoulder pain. *J Orthop Sports Phys Ther* 1996;23:34–8.

- [21] Bamji AN, Erhardt CC, Price TR, Williams PL. The painful shoulder: can consultants agree? *Br J Rheumatol* 1996;35:1172–4.
- [22] Liesdek C, van der Windt D, Koes BW, Bouter LM. Soft tissue disorders of the shoulder. *Physiotherapy* 1997;83:12–7.
- [23] Smidt N, Green S. Is the diagnosis important for the treatment of patients with shoulder complaints? *Lancet* 2003;362:1867–8.
- [24] Bielle A. Epidemiology of shoulder problems. *Baillieres Clin Rheumatol* 1989;3:437–51.
- [25] May S, Littlewood C, Bishop A. Reliability of procedures used in the physical examination of non-specific low back pain: a systematic review. *J Physiother* 2006;52:91–102.
- [26] Simel DL, Rennie D, Bossuyt PM. The STARD statement for reporting diagnostic accuracy studies: application to the history and physical examination. *J Gen Int Med* 2008;23:768–74.
- [27] Khan AY, Preskorn SH, Baker B. Effect of study criteria on recruitment and generalizability of the results. *J Clin Psychopharmacol* 2005;25:271–5.
- [28] Morris R, MacNeela P, Scott A, Treacy P, Hyde A, O'Brien J, et al. Ambiguities and conflicting results: the limitations of the kappa statistic in establishing the interrater reliability of the Irish nursing minimum data set for mental health: a discussion paper. *Int J Nurs Stud* 2008;45:645–7.
- [29] Tooth LR, Ottenbacher KJ. The kappa statistic in rehabilitation research: an examination. *Arch Phys Med Rehabil* 2004;85:1371–6.
- [30] Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther* 2005;85:257–68.
- [31] Ansari-Lari M. Comparison between two tests results, kappa statistic instead of simple overall agreement. *Vet Parasitol* 2005;133:369–70.
- [32] Cyr L, Francis K. Measures of clinical agreement for nominal and categorical data: the kappa coefficient. *Comput Biol Med* 1992;22:239–46.
- [33] Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74.
- [34] Steenbrink F, de Groot JH, Veeger HE, Meskers CG, van de Sande MA, Rozing PM. Pathological muscle activation patterns in patients with massive rotator cuff tears, with and without subacromial anaesthetics. *Man Ther* 2006;11:231–7.
- [35] Aina A, May S. A shoulder derangement. *Man Ther* 2005;10:159–63.
- [36] Quality Management System for Soft Tissue Injuries (QMS-STI). A joint pilot project with rehabilitation providers in Ontario and the Institute for Work & Health. Unpublished working paper; Toronto: Institute for Work & Health; 1998.

Available online at www.sciencedirect.com



ScienceDirect