



Research Article

DOI: 10.36959/987/240

Inter-Observer Reliability of Physical Examination in the Painful Shoulder: Supraspinatus Tendinopathy

Jose Manuel Moreno-Fernandez*, Francisco Martinez-Martinez and Fernando Santonja-Medina

Orthopaedic Surgery and Traumatology Department, Hospital Clínico Universitario Virgen de la Arrixaca, Spain

Abstract

Objectives: Our hypothesis is that there is enough concordance in the implementation and interpretation of the orthopaedic maneuvers by expert explorers. The aim of our study was to analyze the inter-observer reliability of special orthopaedic maneuvers aimed at the physical examination of the supraspinatus tendon.

Setting: Secondary care; referral hospital for the Region de Murcia (Spain) fifth level of care.

Participants: 66 patients, 32 men and 34 women, were explored. The patients included were adults (≥ 18 years), who suffered one-sided omalgia during at least 3 months.

Exclusion criteria were: Bilateral shoulder pain, fractures or previous dislocations, osteoarthritis and advanced retractable capsulitis, previous surgeries, less than 3 months of the last shoulder infiltration, cervical-brachialgias or neurological affectionation, and the existence of obvious deficiency in the collaboration or understanding of the orders effected by the explorer.

Primary and secondary outcome measures: The physical assessment was conducted by two experienced explorers. The drop-arm test, the Jobe empty-can test, the full-can test and shrug sign, were carried out according to the original descriptions. Inter-observer concordance was studied.

Results: The highest levels of inter-explorer concordance were found in the drop-arm test (0.799 PABAK with an 84, 62% of agreement). The full-can test and shrug sign showed a good reliability, while the Jobe test presented a moderate reliability.

Conclusions: The drop-arm test, empty-can test, full-can test and shrug sign met the minimum criteria of percentage of agreement $> 75\%$ and 0.60 inter-observer reliability, by what were considered to be appropriate for their use in physical examination. Therefore, we consider that they are reproducible tests in medical practice for the diagnosis of the pathology of the supraspinatus tendon.

Keywords

Shoulder pain, Supraspinatus tendon, Physical examination, Inter-observer variation

Article Summary

Highlights

- It is important to have physical maneuvers with valid and reliable evidence.
- Drop-arm test inter-explorer reliability was the greatest, with a PABAK of 0.799.
- Empty-can test inter-explorer reliability was moderate like published in literature.
- Full-can test inter-explorer reliability was good, without previous data in literature.
- Shrug sign had an acceptable inter-explorer reliability.

Strengths and limitations of this study

- The strength of the study is the strict adherence to a

standardised study protocol for reproducibility studies, with a training and an overall agreement phase. We selected patients with painful shoulder, therefore this study will be applicable in medical practice.

***Corresponding author:** José Manuel Moreno-Fernandez, Orthopaedic Surgery and Traumatology Department, Hospital Clínico Universitario Virgen de la Arrixaca, Ctra, Madrid-Cartagena, s/n, 30120, El Palmar, Murcia, Spain, Tel: +34-645489502

Accepted: March 11, 2019

Published online: March 13, 2019

Citation: Moreno-Fernandez JM, Martinez-Martinez F, Santonja-Medina F (2019) Inter-Observer Reliability of Physical Examination in the Painful Shoulder: Supraspinatus Tendinopathy. Arch Sports Med 3(1):129-133

Table 1: Orthopaedic special maneuvers for the exploration used in the study of the rotator cuff.

Test	Reference	Interpretation
Drop arm test	[8]	The appearance of pain or claudication of the affected limb was considered as positive
Jobe test or empty can test	[9]	It was positive if pain or weakness appeared with respect to the contralateral limb
Jobe test in external rotation or full can test	[10]	The occurrence of pain was considered positive
Shrug sign	[11]	Considered positive in the elevation of the shoulder when performing arm abduction

- The limitations could be related to measurement data from physical examination, the difference between observers, health care pressure during exploration, the cumulative effect of pain during the exploration, the existence of a washout period and the effect of rest pain.

Introduction

Shoulder pain is a significant cause of morbidity in the general population. In the United Kingdom it is estimated that its prevalence is 16% [1], increasing to 21% [2] in the 70-year-old population. Each year, about 1% of adults over 45-years in the United Kingdom presented a new episode of shoulder pain, of which only 40-50% [3] were consulted for this reason, presenting an incidence of 15 new cases per year for every 1000 patients seen in primary care [4]. Painful shoulder is the third cause of inquiry due to skeletal muscle problems in primary care and the second cause of referral for specialist consultation [1,5].

It is important to have physical maneuvers with valid and reliable evidence [6,7] that complement a correct anamnesis, to obtain an accurate presumption diagnosis. Some orthopaedic maneuvers try to diagnose the affectionation of the rotator cuff of the shoulder through pain provocation maneuvering. These tests can be the drop-arm test [8], the Jobe empty-can test [9], the full-can test [10] and the shrug sign [11]. The validity [12] and reliability [13-17] of these tests has been demonstrated.

Our hypothesis is that there is enough concordance in the implementation and interpretation of the orthopaedic maneuvers by expert explorers for the supraspinatus tendinopathy.

The aim of our study was to analyze the inter-observer reliability of maneuvers aimed at the physical exploration of the painful shoulder due to supraspinatus tendon pathology.

Material and Methods

Selection of expert explorers

The physical examination of patients was conducted by two Orthopaedic Surgery and Traumatology specialists with over 20 years of experience.

Both received at least 4 training sessions specific to the selected tests. Healthy volunteers were used in a first session to agree on the criteria of execution and interpretation of tests. In subsequent quarterly sessions suggestions and issues that arose were exposed and modifications and necessary

clarifications were provided.

Selection of patients

The patients studied were treated from January 2013 to May 2014 by explorers in the consultations of Orthopaedic Surgery and Traumatology of the Hospital Clinico Universitario Virgen de la Arrixaca (Murcia, Spain), being included in the study consecutively according to the criteria laid down.

The patients included were men and women, adults (≥ 18 years), who suffered one-sided omalgia during at least 3 months.

Exclusion criteria were: Patients with bilateral shoulder pain, fractures, and previous dislocations that could alter the dynamics of the shoulder (middle or distal third of proximal humerus, clavicle and scapula), osteoarthritis and advanced retractable capsulitis, previous surgeries with the last shoulder infiltration of less than 3 months, symptoms of cervical brachialgias or neurological affectionation, and the existence of obvious deficiency in the collaboration or understanding of the orders effected by the explorer.

Evaluation of the shoulder

We selected physical tests which have been proposed as more useful and reproducible in daily clinical practice, aimed at the assessment of the painful shoulder due to tendon pathology of supraspinatus (Table 1).

Statistical analysis

For the statistical calculations the Statistical Package for the Social Sciences (SPSS) version 19.0 (IBM Company) was used.

We calculated the concordance between observers of the tests and physical signs were assessed. For the qualitative variables the percentage of accord between the explorers and the corrected kappa coefficient [18]. We calculated the adjusted kappa index or PABAK (prevalence and bias adjusted kappa) [19] to take into account the degree of disaccord and differences between the proportions of positive and negative outcome that negatively affect the kappa coefficient [20-22].

To find out what were the limits of prevalence or bias affecting the overall kappa value, we calculated the rate of prevalence (PI) and bias (BI) for each variable index [19]. When the PI was high, we used the PABAK value to interpret the results of reliability. For our study, we determined a value of arbitrary cut of less than -0.5 PI, or greater than

0.5, to base our interpretation on PABAK values instead of the kappa value. The kappa and PABAK values for the inter-explorer reliability are interpreted in accordance with the recommendations of Landis and Kock [23]: Less than 0.20 poor; 0.21 to 0.40 regular; 0.41-0.60, moderate; 0.61 to 0.80, good; and 0.81 to 1 very good.

In our study, following the recommendations of Cadogan [17], we established minimum matching criteria so that a test may be considered appropriate for use in physical examination. When we used the absolute percentage of agreement, the minimum value that was established as acceptable was 75% [24,25]. Regarding the inter-observer reliability, the minimum value that was accepted was 0.60 [23,26].

Results

We evaluated 66 patients with shoulder pain, 32 men

Table 2: Descriptive data of the qualitative variables of the anamnesis.

	n	%
Sex		
Male	32	48.48
Female	34	51.51
Laterality		
Left	17	25.76
Right	49	74.24
Dominance		
Left	4	6.06
Right	62	93.94

and 34 women, with an average age of 56 years (range 23-81 years). The average evolution time of painful symptoms was 13 months (Tables 2 and Table 3).

The values of PI, BI, corrected kappa, PABAK and the percentage of agreement with the special maneuvers of exploration are represented in Table 4. Inter-explorer concordance in this section is highly variable. The PI exceeded limits in all cases, so the PABAK values are representative of these tests. The PABAK values oscillated between 0.505-0.799 and indicated moderate-good concordance between explorers. The percentage of agreement was 80% to 88%.

Discussion

May [27] presented a systematic review of six studies of reliability on procedures of physical examination of painful shoulder, including 17 high quality studies, which showed conflicting results and most with values below acceptable levels of reliability. Nomden [14] presented a reliability study of 23 tests of shoulder girdle physical examinations made by physiotherapists. The tests were not standardized or backed by bibliography. They concluded that around 50% of the tests used did not meet the statistical criteria for acceptable reliability.

Cadogan [17] conducted a study with 40 patients and reported a good concordance between examiners (PABAK 0.67) in the drop-arm test. Our reliability in this case was greater, with a PABAK of 0.799 and a percentage of agreement of 84, 62%.

The empty-can test in our study showed a moderate reliability with a PABAK of 0.505. Ostor [13] studied 159 shoulders by a rheumatologist expert, a rheumatologist without experience in this field and a nurse, who were

Table 3: Descriptive data of the quantitative variables of the anamnesis.

	Medium	Range			p
Evolution time (months)	13	3		132	
Age (years)	55.98	23	-	81	
- Males	55.34	25	-	81	0.732
- Females	56.59	23	-	78	
DASH	49.79				
- Males	45.48	12.5	-	81.67	0.084
- Females	54.10	19.16	-	89.17	
EVA					
- Exploration 1	6.96				0.831
- Exploration 2	6.65				

Table 4: Reliability of special maneuvers of exploration data.

Special maneuvers	% agreement	PI	BI	Kappa corrected	PABAK
Drop-arm test	84.62	0.69	0.15	0.63	0.799
Jobe	80.00	0.64	0.16	0.196	0.505
Jobe in ER	88.00	0.76	0.12	0.452	0.761
Shrug sign	80.00	0.60	0.20	0.611	0.735

Table 5: Review of the reliability of special orthopaedic maneuvers of exploration of the shoulder.

Maneuvers	Reliability	Reference	Interpretation
Drop-arm test	0.47 (0.28-0.66)	[30]	Moderate
	0.467 (0.275-0.659)	[13]	Moderate
	0.67	[17]	Good
	0.799	Our study	Good
Test of Jobe	0.47 (0.22-0.72)	[15]	Moderate
	0.46 (0.44-0.49)	[13]	Moderate
	0.505	Our study	Moderate
Test of Jobe in ER	No previous study	-	-
	0.761	Our study	Good
Shrug Sign	0.833	[29]	Very good
	0.735	Our study	Good

instructed by a phase of training. They presented inter-observer moderate reliability for the Jobe test similar to our study, stressing the importance of sessions of updating and training of specialists for the diagnosis of painful shoulder management. Michener [15] studied 55 patients with painful shoulders that were explored by an orthopaedic surgeon and an expert physiotherapist using a combination of 5 physical tests for the diagnosis of subacromial impingement syndrome. For the test of Jobe, a moderate reliability, similar to our study, was shown.

Vind [16] studied shoulders in healthy athletes who were overstraining above their heads and reported a reliability of 0.9 with a percentage of agreement of 95%. They included a training phase with 10 healthy volunteers, a phase of a global agreement with 20 players of handball and a phase study of 50% prevalence with 44 subjects, which may explain the greater concordance provided. Palmer [28] did not mention in his Protocol of Exploration of Southampton the Jobe test, classically described and used in our study, they presented it as the presence of pain in the shoulder to resisted abduction. The data presented in this study for this test were superior to ours, with a reliability of 0.81 and a percentage of agreement of 94%.

The full-can test in our study showed good reliability with a PABAK 0.761 and a percentage of agreement of 88%. We believe that this contribution is relevant, since we did not find data published in the literature consulted.

The shrug sign in our study showed a good reliability of 0.735, with a percentage of agreement of 80%. Jia [29] studied the inter-observer reliability of shrug sign, showing a very good reliability with a kappa of 0.833.

The Table 5 shows a review of the reliability of special orthopaedic maneuvers used in our study.

We believe that there are several limitations that influenced this moderate-good overall reliability. These limitations could be related to measurement data from physical examination, the difference between observers, health care pressure during exploration, the cumulative effect of pain during the exploration, the existence of a washout period and the effect of rest pain.

Given that in the consulted literature we did not find any studies that analyze the reliability of the full-can test or set of tests of the supraspinatus, we believe that our work provides relevant information about the semiology of the supraspinatus tendon pathology.

Conclusions

The authors conclude that the drop-arm test, empty-can test, full-can test and shrug sign have an acceptable inter-explorer reliability. Therefore, we consider that these tests are reproducible in clinical practice for the diagnosis of the pathology of the supraspinatus tendon.

Acknowledgements

The authors thank the selfless and voluntary participation of all patients included in this study.

Contributorship Statement

JMMF: Substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work; AND Drafting the work or revising it critically for important intellectual content; AND Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

FMM: Substantial contributions to the conception or design of the work; or the acquisition, AND Final approval of the version to be published; AND Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

FSM: Substantial contributions to the conception or design of the work; or the acquisition; AND Drafting the work or revising it critically for important intellectual content; AND Final approval of the version to be published; AND Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Competing Interests

All authors have completed the ICMJE uniform disclosure

form at www.icmje.org/coi_disclosure.pdf and declare: No support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Funding

There are no funders to report for this study.

Data Sharing Statement

No additional data available.

References

1. Urwin M, Symmons D, Allison T, et al. (1998) Estimating the burden of musculoskeletal disorders in the community: The comparative prevalence of symptoms at different anatomical sites, and the relation to social deprivation. *Ann Rheum Dis* 57: 649-655.
2. Chard MD, Hazleman R, Hazleman BL, et al. (1991) Shoulder disorders in the elderly: A community survey. *Arthritis Rheum* 34: 766-769.
3. Bongers PM (2001) The cost of shoulder pain at work. *BMJ* 322: 64-65.
4. Van der Windt DA, Koes BW, de Jong BA, et al. (1995) Shoulder disorders in general practice: Incidence, patient characteristics, and management. *Ann Rheum Dis* 54: 959-964.
5. Butcher JD, Zukowski CW, Brannen SJ, et al. (1996) Patient profile, referral sources, and consultant utilization in a primary care sports medicine clinic. *J Fam Pract* 43: 556-560.
6. Krebs DE (1987) Measurement theory. *Phys Ther* 67: 1834-1839.
7. Fritz JM, Wainner RS (2001) Examining diagnostic tests: An evidence-based perspective. *Phys Ther* 81: 1546-1564.
8. Codman EA (1934) *The shoulder; rupture of the supraspinatus tendon and other lesions in or about the subacromial bursa*, Boston.
9. Jobe FW, Jobe CM (1983) Painful athletic injuries of the shoulder. *Clin Orthop Relat Res* 173: 117-124.
10. Kelly BT, Kadrmaz WR, Speer KP (1996) The manual muscle examination for rotator cuff strength. An electromyographic investigation. *Am J Sports Med* 24: 581-588.
11. Ktrak PH (1990) Shoulder shrug--a prognostic sign for recovery of hand movement after stroke. *Med J Aust* 152: 297-301.
12. Hegedus EJ, Goode A, Campbell S, et al. (2008) Physical examination tests of the shoulder: A systematic review with meta-analysis of individual tests. *Br J Sports Med* 42: 80-92.
13. Ostor AJ, Richards CA, Prevost AT, et al. (2004) Interrater reproducibility of clinical tests for rotator cuff lesions. *Ann Rheum Dis* 63: 1288-1292.
14. Nomden JG, Slagers AJ, Bergman GJ, et al. (2009) Interobserver reliability of physical examination of shoulder girdle. *Man Ther* 14: 152-159.
15. Michener LA, Walsworth MK, Doukas WC, et al. (2009) Reliability and diagnostic accuracy of 5 physical examination tests and combination of tests for subacromial impingement. *Arch Phys Med Rehabil* 90: 1898-1903.
16. Vind M, Bogh SB, Larsen CM, et al. (2011) Inter-examiner reproducibility of clinical tests and criteria used to identify subacromial impingement syndrome. *BMJ Open* 1: e000042.
17. Cadogan A, Laslett M, Hing W, et al. (2011) Interexaminer reliability of orthopaedic special tests used in the assessment of shoulder pain. *Man Ther* 16: 131-135.
18. Cohen J (1960) A coefficient of agreement for nominal scales. *Educ Psychol Meas* 20: 37-46.
19. Byrt T, Bishop J, Carlin JB (1993) Bias, prevalence and kappa. *J Clin Epidemiol* 46: 423-429.
20. Feinstein AR, Cicchetti DV (1990) High agreement but low Kappa: I. The problems of two paradoxes. *J Clin Epidemiol* 43: 543-549.
21. Rigby AS (2000) Statistical methods in epidemiology. V. Towards an understanding of the kappa coefficient. *Disabil Rehabil* 22: 339-344.
22. Shankar V, Bangdiwala SI (2008) Behaviour of agreement measures in the presence of zero cells and biased marginal distributions. *J Appl Statistics* 35: 445-464.
23. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. *Biometrics* 33: 159-174.
24. Hartmann DP (1977) Considerations in the choice of interobserver reliability measures. *J Appl Behav Anal* 10: 103-116.
25. Stemler SE (2004) A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Practical Assessment, Research & Evaluation* 9: 1-11.
26. Altman DG (1991) *Practical statistics for medical research*, Chapman & Hall, London.
27. May S, Chance-Larsen K, Littlewood C, et al. (2010) Reliability of physical examination tests used in the assessment of patients with shoulder problems: A systematic review. *Physiotherapy* 96: 179-190.
28. Palmer K, Walker-Bone K, Linaker C, et al. (2000) The Southampton examination schedule for the diagnosis of musculoskeletal disorders of the upper limb. *Ann Rheum Dis* 59: 5-11.
29. Jia X, Ji JH, Petersen SA, et al. (2008) Clinical evaluation of the shoulder shrug sign. *Clin Orthop Relat Res* 466: 2813-2819.
30. Park HB, Yokota A, Gill HS, et al. (2005) Diagnostic accuracy of clinical tests for the different degrees of subacromial impingement syndrome. *J Bone Joint Surg Am* 87: 1446-1455.

DOI: 10.36959/987/240