

BMJ Open

BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (<http://bmjopen.bmj.com>).

If you have any questions on BMJ Open's open peer review process please email editorial.bmjopen@bmj.com

BMJ Open

Inter-tester reliability of clinical shoulder instability and laxity tests in subjects with and without self-reported shoulder problems

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2017-018472
Article Type:	Research
Date Submitted by the Author:	30-Jun-2017
Complete List of Authors:	Eshoj, Henrik; Syddansk Universitet Det Sundhedsvidenskabelige Fakultet, Department of Sports Science and Clinical Biomechanics; Odense Universitetshospital, Quality of Life Research Center, Department of Haematology Ingwersen, Kim Gordon; Hospital Lillebaelt - Vejle Hospital, Vejle, Denmark, Research Unit at Department of Physical- and Occupational Therapy,; Syddansk Universitet Det Sundhedsvidenskabelige Fakultet, Department of Sports Science and Clinical Biomechanics Larsen, Camilla; Research Unit for Musculoskeletal Function and Physiotherapy, Institute of Sports Science and Clinical Biomechanics Kjær, Birgitte; Bispebjerg and Frederiksberg University Hospital, Copenhagen, Denmark , Research Unit at Department of Physical- and Occupational Therapy, ; Syddansk Universitet Det Sundhedsvidenskabelige Fakultet, Department of Sports Science and Clinical Biomechanics Juul-Kristensen, Birgit; Research Unit for Musculoskeletal Function and Physiotherapy, Institute of Sports Science and Clinical Biomechanics
Primary Subject Heading:	Evidence based practice
Secondary Subject Heading:	Sports and exercise medicine
Keywords:	Reliability, Shoulder < ORTHOPAEDIC & TRAUMA SURGERY, Instability, Laxity, Clinical test

SCHOLARONE™
Manuscripts

1
2
3
4 1 **Inter-tester reliability of clinical shoulder instability and laxity tests in subjects**
5
6
7 2 **with and without self-reported shoulder problems**
8
9

3

10
11 4 Henrik Eshoj¹, PhD^{a,b} Kim Gordon Ingwersen, PhD^{a,c}, Camilla Marie Larsen, PhD^{a,d} Birgitte Hougs
12
13 5 Kjær, Msc^{a,e}, Birgit Juul-Kristensen, PhD^{a,f}
14
15
16
17 6

7 **Affiliations**

8 ^a Department of Sports Science and Clinical Biomechanics, University of Southern Denmark,
9 Odense, Denmark

10 ^b Quality of Life Research Center, Department of Haematology, Odense University Hospital,
11 Odense, Denmark

12 ^c Research Unit at Department of Physical- and Occupational Therapy, Hospital Lillebaelt - Vejle
13 Hospital, Vejle, Denmark

14 ^d Health Sciences Research Centre, University College Lillebaelt, Odense, Denmark

15 ^e Research Unit at Department of Physical- and Occupational Therapy, Bispebjerg and
16 Frederiksberg University Hospital, Copenhagen, Denmark

17 ^f Institute of Occupational Therapy, Physiotherapy and Radiography, Department of Health
18 Sciences, Bergen University College, Norway
19

20 **Correspondence:**

21 Birgit Juul-Kristensen

22 Department of Sports Science and Clinical Biomechanics

23 University of Southern Denmark, Campusvej 55, 5230 Odense M, DK

24 Tel.: +45 2942 3335

25 Email: bjjuul-kristensen@health.sdu.dk

26 **Word counts: 2866**
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4 28 **Abstract**

5
6
7 29 **Objective:** Firstly, to investigate the inter-tester reliability of clinical shoulder instability and laxity
8
9 30 tests, and secondly, to describe the mutual dependency of each of the individual tests for
10
11 31 identifying self-reported shoulder problems.
12
13

14 32
15
16 33 **Method:** A standardized protocol for conducting reliability studies was used to test the inter-
17
18 34 tester reliability of the six clinical shoulder instability and laxity tests; apprehension, relocation,
19
20
21 35 surprise, load-and-shift, sulcus sign and Gagey. Cohens kappa (k) with 95% confidence intervals
22
23 36 (CI) besides Prevalence-Adjusted-Bias-Adjusted-Kappa (PABAK), accounting for insufficient
24
25 37 prevalence and bias, were computed to establish the inter-tester reliability and mutual
26
27
28 38 dependency.
29
30

31 39
32
33 40 **Results:** Forty individuals (13 with instability and laxity related shoulder problems and 27 shoulder
34
35 41 healthy individuals) aged 18-60 were included. Fair (relocation), moderate (load-and-shift, sulcus
36
37 42 sign) and substantial (apprehension, surprise, Gagey) inter-tester reliability were observed across
38
39
40 43 tests (k 0.39-0.73; 95% CI: 0.00-1.00). PABAK improved reliability across tests, resulting in
41
42 44 substantial to almost perfect inter-tester reliability for the apprehension, surprise, load-and-shift
43
44 45 and Gagey tests (k 0.65-0.90). Mutual dependencies between each test and self-reported shoulder
45
46 46 problem showed apprehension, relocation and surprise to be the most often used tests to
47
48 47 characterize such musculoskeletal shoulder conditions.
49
50

51 48
52
53 49 **Conclusions:** Four tests (apprehension, surprise, load-and-shift and Gagey) out of six, were
54
55 50 considered inter-tester reliable for clinical use, while relocation and sulcus sign tests need further
56
57
58
59
60

1
2
3
4 51 standardization before acceptable evidence. Furthermore, the validity of the tests for instability
5
6 52 and laxity needs to be studied.
7
8

9 53

10 11 54 Article summary section

12 55

13 56 Strengths and limitations of this study

14 57

- 15 58 • The strength of the study is the use of a three-phased standardized study protocol
- 16 59 • Presentation of raw findings increases transparency and interpretation of study findings
- 17 60 • No valid gold standard for including shoulder instability and laxity subjects was used
- 18 61 • A 50/50 prevalence of positive and negative tests for all six tests was not accomplished

19
20
21
22
23
24
25
26
27
28 62

29 63 Keywords

30 64 Reliability, Shoulder instability, laxity, clinical tests

31
32
33
34
35 65

36 66 Acknowledgements

37 67 We would like to thank Physiotherapists Rasmus Fitzner, Pernille Madsen and Jacob Hansen from
38
39 68 Metropolitan University College, Copenhagen, Denmark for recruitment and testing of study
40
41
42 69 participants. Furthermore, a special thanks to Bispebjerg Frederiksberg University Hospital,
43
44
45 70 Copenhagen, Denmark for providing facilities for data collection.
46
47

48 71

49
50 72
51
52
53
54
55
56
57
58
59
60

73 Introduction

74 Shoulder complaints, affecting shoulder-related quality of life (QoL), are frequent and may be
75 caused by shoulder instability and/or laxity¹ due to traumatic or non-traumatic injuries to the
76 shoulder joint.² The traumatic shoulder instability is mainly prompted by a high impact injury
77 during sports participation, resulting in a shoulder dislocation, predominantly in anterior
78 direction.³ The non-traumatic shoulder instability is usually related to repetitive overhead
79 activities and/or patients with generalised joint hypermobility or glenohumeral hyperlaxity, often
80 referred to as multidirectional shoulder instability.^{2,4,5}

81 Irrespectively of aetiology, shoulder instability and laxity is often accompanied by a variety of
82 symptoms including shoulder discomfort, pain besides glenohumeral subluxations and/or
83 repeated dislocations.^{6,7,8} Clinically, shoulder instability and laxity, is diagnosed and verified by a
84 group of shoulder pain and instability provoking/relieving tests, supplemented by shoulder laxity
85 tests.^{9,10} The former tests usually include the anterior shoulder instability and laxity tests;
86 apprehension, relocation and surprise, and the laxity tests consisting of the load-and-shift, sulcus
87 sign and Gagey tests.^{11,12,13} An ongoing discussion is the use of pain as diagnostic criterion in
88 diagnosing anterior shoulder instability with the shoulder-provoking tests apprehension,
89 relocation and surprise. In one way, it may be a confounding factor, since pain has shown to be
90 less predictive and reliable as a diagnostic criterion.¹⁴ On the contrary though, others have
91 suggested that unrecognized and underlying glenohumeral instability may lead to repetitive
92 microtrauma and painful shoulder conditions,^{15,16} justifying pain as diagnostic criterion when
93 testing for anterior shoulder instability.

94 Nonetheless, symptoms may become chronic, and lead to reduced work and sports
95 capability,^{17,18,19} and with exercise-based management as the most often recommended first-

Inter-tester reliability of clinical shoulder instability and laxity tests

1
2
3
4 96 choice treatment.^{20,21} Hence, early diagnosis using reliable and accurate clinical tests to guide
5
6 97 focused treatment is essential. Few studies though, have investigated the reliability of clinical
7
8 98 shoulder instability and laxity tests showing large variations in reliability and with limited
9
10 99 methodological quality, hampering interpretation and comparison with other studies.^{14,22,23}
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

100

101 Therefore, the objective of this study was to investigate the inter-tester reliability of commonly
102 used clinical shoulder instability and laxity tests in a group of sports-active individuals with and
103 without self-reported shoulder problems.

104

105

106 **Materials and methods**

107 **Study design**

108 An inter-tester reliability study was conducted involving two physiotherapists as inter-tester
109 examiners. A third physiotherapist (study coordinator), not involved in the actual inter-tester
110 reliability study, managed all practical aspects during the study period. The Guidelines for
111 Reporting Reliability and Agreement Studies (GRASS) were followed.²⁴ A standardized protocol for
112 reliability studies, consisting of three phases: preparation and training of clinical tests, overall
113 agreement, and study phase (the actual reliability study) were applied.²⁵ Two early career
114 physiotherapists with six months clinical experience were involved in the inter-tester reliability
115 study. A test-protocol describing each clinical test was developed and subsequently used by the
116 two testers to practice all tests in order to reach uniformity and mutual agreement in performing
117 and interpreting each test. In the overall agreement phase, the two testers examined 19
118 individuals (eight shoulder cases and 11 shoulder healthy). The two testers were mutually blinded
119 to the health status of the individuals (shoulder cases vs. shoulder healthy) and also to each
120 other's test results. Before proceeding to the final study phase, the two testers needed an overall
121 agreement of at least 80% based on findings from the six clinical shoulder tests.²⁵ In the actual
122 inter-tester reliability study phase, the two testers examined a new group of shoulder cases and
123 shoulder healthy individuals with the six clinical shoulder tests. The procedure was the same as in
124 the agreement phase, meaning that testers were blinded to the health status of the individuals
125 and each other's test results.

126

127

128

1
2
3
4 129 **Study subjects**
5

6 130 A sample size of at least 40 individuals was targeted based on recommendations for performing
7
8 131 clinical reliability studies.²⁵ Sixty-five (women and men (aged 18-60 years)) were recruited and
9
10 132 screened for eligibility from Metropolitan University College, Copenhagen, and Bispebjerg
11
12 133 Frederiksberg University Hospital, Copenhagen resulting in an included number of 13 individuals
13
14 134 with instability and/or laxity related shoulder problems (hereinafter referred to as cases) vs. 27
15
16 135 shoulder healthy individuals, respectively.
17
18 136 Cases answering yes to at least one of two questions (*'Do you have a sense of shoulder instability?'*
19
20 137 and *'Have you ever had a shoulder injury?'*) were eligible for a clinical shoulder examination
21
22 138 performed by the study coordinator. Cases were then included if they present with at least one
23
24 139 positive clinical shoulder test out of the following; apprehension, relocation, surprise, load-and-
25
26 140 shift, sulcus sign or Gagey. The shoulder healthy individuals were included if they present with no
27
28 141 self-reported shoulder pathology or complaints. In general, any individuals with prior shoulder
29
30 142 surgery were excluded. In the actual study phase, individuals completed a short questionnaire
31
32 143 with basic demographic details (age, gender, weight, height), in addition to the following: pain
33
34 144 level during rest and activity (numeric pain rating scale, NPRS),²⁶ shoulder injury ever (yes/no),
35
36 145 subjective shoulder instability (yes/no) and sports-related activity (hours/week). Further, all
37
38 146 individuals filled in the patient-reported Western Ontario Shoulder Instability (WOSI)
39
40 147 questionnaire designed to measure shoulder function and quality of life in patients with shoulder
41
42 148 instability and laxity symptoms.²⁷ The study was exempted for notification to the Danish Health
43
44 149 Research Study Board due to the non-invasive/-non-treating study design. However, oral and
45
46 150 written consent was provided from all individuals and, ethical guidelines were followed according
47
48 151 to the Helsinki declaration.²⁸
49
50
51
52
53
54
55
56
57
58
59
60

Inter-tester reliability of clinical shoulder instability and laxity tests

152 **Clinical tests**

153 The clinical shoulder tests consisted of three shoulder joint-provoking tests for anterior shoulder
154 instability (apprehension, relocation and surprise) besides three shoulder laxity tests (load and
155 shift, sulcus sign and Gagey) (Table 1).^{11,13,14,22,23,29}

156 The apprehension and surprise tests were positive if glenohumeral apprehension and/or pain
157 were evoked during testing whereas relieve of symptoms with the relocation test was regarded as
158 a positive test. The load and shift test was rated on a four-point scale ranging from 0-3 (best to
159 worst; 0= little glenohumeral movement; 3=humeral head moves beyond the glenoid rim and
160 remains dislocated).¹² Also, to enhance mutual agreement between testers when performing the
161 load-and-shift test, only the direction (anterior vs. posterior) with most glenohumeral head
162 translation was rated. Sulcus sign was objectively measured in centimeter (continuous scale) by
163 use of a small ruler according to previously used grading scales as follows; I (<1 cm translation), II
164 (1 to 2.0 cm) or III (>2.0 cm).²⁹ Finally, Gagey test was rated as positive with passive abduction
165 above 105 degrees.¹³

167 **Statistics**

168 Demographics and descriptive data were tested for normality by visual inspection of histograms
169 and Shapiro-Wilk's test. Group differences (shoulder cases vs. shoulder healthy) were tested by
170 Fisher's Exact test for categorical variables, whereas student's t-test and Mann-Whitney U-test
171 was used for parametric, respectively non-parametric distributed data.

172 Apprehension, relocation, surprise and Gagey test were dichotomous variables whereas the load
173 and shift and sulcus sign test were dichotomized to also allow for nominal statistics. Thus, load-
174 and-shift was rated positive when scored 2 or 3, while for sulcus sign a positive rating was equal to



Inter-tester reliability of clinical shoulder instability and laxity tests

1
2
3
4 175 measurements exceeding one centimeter.²⁹ For transparency, data from each test is presented by
5
6 176 2 x 2 contingency tables besides the use of McNemar's test for significant between-tester
7
8
9 177 differences. Furthermore, observed and expected agreements are presented along with
10
11 178 prevalence and bias³⁰ indexes. Reliability was evaluated with the use of Cohen's kappa (*k*)
12
13 179 coefficients including 95% confidence intervals (CI).²⁵ Also, since kappa is sensitive to imbalances
14
15 180 in prevalence and bias (e.g. the number of positive and negative tests not close to 50%) a
16
17
18 181 Prevalence-Adjusted-Bias-Adjusted Kappa (PABAK) was calculated.^{30,31} By definition, PABAK
19
20 182 reflects the ideal situation, thereby accounting for variation of prevalence and bias between
21
22
23 183 testers (as presented in the "real" world).³² The relationship between the individual tests and the
24
25 184 classification (mutual dependency) by self-reported shoulder problems was tested by Cohen's
26
27
28 185 kappa (*k*) coefficients and the characterization of the groups was tested with Fischer's exact tests.
29
30 186 The classification system proposed by Landis and Koch was used to interpret reliability as follows:
31
32 187 0.00-0.20 (Slight); 0.21-0.40 (Fair); 0.41-0.60 (Moderate); 0.61-0.80 (Substantial) and 0.81-1.00
33
34 188 (Almost perfect).³³
35
36
37 189 Statistical Package for the Social Sciences (SPSS inc., Chicago, IL, USA), version 22, was used for all
38
39
40 190 statistical analyses, with p-value of <0.05 interpreted as significant.
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Inter-tester reliability of clinical shoulder instability and laxity tests

Table 1. Performance and evaluation of the clinical shoulder instability and laxity tests**Verbal introduction:**

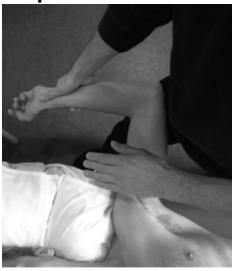
1. I am going to perform six clinical shoulder tests on you
2. I will ask if you experience any symptoms (apprehension and/or pain) during the three first tests.
3. I will guide you through each test

Clinical tests	Description	Placing of hands, etc.	Evaluation
Apprehension 	<p>Individuals placed supine with the shoulder being tested close to the edge of the examination table.</p> <p>Shoulder positioned in 90° of abduction, elbow flexed to 90°.</p> <p>Examiner moves the shoulder into maximal external rotation.</p>	<p>One hand around the wrist of the individual with the other hand gently placed in front of the shoulder.</p> <p>Elbow supported at the examiners thigh.</p>	<p>Subjective or objective presence of apprehension and/or pain?</p> <p>Rated as either positive or negative.</p> <p>(Nominal, dichotome data)</p>
Relocation 	<p>From the end position of the apprehension test the humeral head is gently forced posteriorly</p>	<p>Examiners fifth finger placed close to the lateral part of the acromion with the wrist positioned anteriorly of the humeral head.</p>	<p>Diminish of apprehension symptoms and/or pain?</p> <p>Rated as either positive or negative.</p> <p>(Nominal, dichotome data)</p>

Inter-tester reliability of clinical shoulder instability and laxity tests

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

Surprise



From end position of the relocation test the posteriorly directed force at the humeral head is quickly removed.

Removal of examiners wrist from the anterior part of the shoulder.

Subjective or objective reproduction of apprehension symptoms and/or pain?
Rated as either positive or negative.

(Nominal, dichotome data)

Load-and-shift

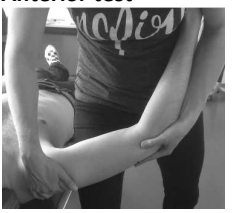
Individual placed supine with scapula resting at the examination table.
Humeral head is loaded gently into the glenoid through axial pressure at the elbow.

Examiners one hand placed at the olecranon with the individual's hand positioned between the examiners torso and elbow.

Humeral head movement evaluated by the use of a four-level laxity scale.

0 = little to almost no movement

Anterior test



Anterior load-and-shift:
Shoulder positioned in the scapular plane in 90° of abduction with elbow flexed.
Humeral head gently shifted in anterior direction.

Anterior load-and-shift: Examiners hand placed on top of the shoulder with the fingers on the backside of the glenohumeral head to move it anteriorly.

1 = humeral head moves up onto the glenoid

2 = humeral head moves beyond the glenoid, but relocates spontaneously once pressure is released

Posterior test



Posterior load-and-shift: Shoulder positioned in the scapular plane in 20° of abduction with elbow flexed.
Humeral head gently shifted in posterior direction.

Posterior load-and-shift: Examiners wrist placed at the anterior part of the humeral head to move it posteriorly.

3 = humeral head moves beyond the glenoid and remains dislocated

Rated as positive when scored 2 or 3.

Inter-tester reliability of clinical shoulder instability and laxity tests

(Nominal, dichotome data)

Sulcus sign



Individual sitting upright.

Examiner pulls the distal part of the humerus in a caudal direction.

Distance from the top of the humeral head and the acromion is measured with a ruler.

One hand placed above the epicondyles of humerus.

Other hand is used to measure the subacromial distance with a ruler

Rated as positive with measurements exceeding 1 centimeter.

(Nominal, dichotome data)

Gagey



Individual sitting upright.

The shoulder girdle is gently depressed with the arm passively moved into end range in horizontal abduction.

A mirror in front of the individual is used to evaluate shoulder abduction angle.

Forearm placed on top of the shoulder girdle with the other hand placed around the elbow joint.

Rated as positive with abduction exceeding 105 degrees.

(Nominal, dichotome data)

191 **Results**

192 Characteristics of the participating individuals are presented in Table 2. Demographics
 193 showed no difference between the shoulder cases (n=13) and shoulder healthy (n=27)
 194 individuals. Further, both groups (92 and 74%; p=0.18) were relatively active with a
 195 weekly participation in sports-related activity for more than four hours per week.
 196 However, as expected due to the design, shoulder cases had significantly higher pain
 197 during activity (4.23 vs. 1.44; p=0.02), higher frequency of shoulder injury ever (62% vs.
 198 <1%; p<0.001), higher subjective shoulder instability (69 vs. 11%; p<0.001) and worse
 199 total WOSI score (506 vs. 136; p=0.001) (Table 2).

201 **Table 2.** Study phase. Participant characteristics.

	Shoulder cases (n=13)	Shoulder healthy (n=27)	P-value
Sex (women/men)	8/5	21/6	0.28
Age (years) mean (SD)	28 (9)	29 (7)	0.72
Weight (kg)* mean (SD)	71.0 (12.8)	74.9 (23.4)	0.59
Height (cm) mean (SD)	174.0 (8.6)	173.4 (7.9)	0.82
Pain, rest (NRS 0-10) mean (SD)	1.08 (1.44)	0.41 (1.15)	0.12
Pain, activity (NRS 0-10) mean (SD)	4.23 (2.92)	1.44 (2.12)	<0.05
Shoulder injury ever, n (%)	8 (62)	1 (<1)	<0.001
Subjective shoulder instability, n (%)	9 (69)	3 (11)	<0.001
Sports-related activity (>4 hours/week), n (%)	12 (92)	20 (74)	0.18
WOSI domains, mean (SD)			
- Physical symptoms (0-1000)	225 (165)	60 (78)	<0.05
- Sports, recreation, work (0-400)	103 (93)	24 (47)	<0.05
- Lifestyle (0-400)	58 (57)	13 (21)	<0.05
- Emotions (0-300)	121 (94)	39 (49)	<0.05
WOSI total score (0-2100) mean (SD)	506 (362)	136 (174)	<0.001

SD Standard deviation; kg kilo; * significance level p < 0.05; cm centimeter NRS Numeric Rating Scale

202

203

204

Inter-tester reliability of clinical shoulder instability and laxity tests

205 Prevalence of positive tests was especially low for the load-and-shift test (table 3), and
 206 significant between-tester differences were found for relocation and sulcus sign tests
 207 ($p=0.021$) (not shown in tables).

208

209 **Table 3.** Contingency tables with findings from tester A and B.

Apprehension		A	
		Yes	No
B	Yes	14	4
	No	3	19

Relocation		A	
		Yes	No
B	Yes	6	2
	No	8	24

Surprise		A	
		Yes	No
B	Yes	14	4
	No	3	19

Load-and-shift		A	
		Yes	No
B	Yes	1	0
	No	2	37

Sulcus		A	
		Yes	No
B	Yes	7	1
	No	9	23

Gagey		A	
		Yes	No
B	Yes	8	3
	No	1	28

210

211

212 Reliability varied between k : 0.39-0.73 (95% CI: 0.00-1.00), indicating fair (relocation; k
 213 0.39), moderate (load-and-shift, sulcus sign; k 0.43 and 0.48) and substantial
 214 (apprehension, surprise, Gagey; k 0.65-0.73) reliability (Table 4). The prevalence index
 215 of all six tests ranged from 0.05-0.44, (lowest for load-and-shift, relocation and sulcus;
 216 0.05, 0.28 and 0.30), whereas the bias index ranged from and 0.03-0.20 (highest for
 217 relocation and sulcus). PABAK improved reliability for relocation, load-and-shift, sulcus
 218 sign and Gagey test, now corresponding to moderate (relocation and sulcus sign; k
 219 0.50), substantial (Gagey; k : 0.80) and almost perfect (load and shift; k : 0.90) reliability
 220 (Table 4).

221

222

Inter-tester reliability of clinical shoulder instability and laxity tests

Table 4. Study phase. Reliability of six clinical shoulder instability and laxity tests.

	Observed agreement	Expected agreement	Prevalence index	Bias index	Kappa (95% CI)	PABAK
Apprehension	0.83	0.51	0.44	0.03	0.65 (0.38; 0.85)	0.65
Relocation*	0.75	0.59	0.28	0.15	0.39 (0.07; 0.68)	0.50
Surprise	0.83	0.51	0.44	0.03	0.65 (0.38; 0.85)	0.65
Load and Shift	0.95	0.90	0.05	0.05	0.48 (0.00; 1.00)	0.90
Sulcus sign*	0.75	0.56	0.30	0.20	0.43 (0.17; 0.72)	0.50
Gagey	0.90	0.62	0.40	0.05	0.73 (0.46; 0.94)	0.80

*Significant inter-tester differences, CI Confidence intervals, PABAK prevalence-and-bias-adjusted kappa

The *k* values for mutual dependency indicate that apprehension, relocation and surprise tests for both examiners were the most frequently used tests for characterizing self-reported shoulder problems (Table 5). This was further confirmed by the significant group difference in the presence of positive tests.

Table 5. Kappa statistics for mutual dependency of the individual tests and self-reported shoulder problems.

	Observed agreement	Expected agreement	Prevalence index	Kappa	p-value (cases/controls)
Apprehension					
Examiner A	0.75	0.53	0.38	0.47	0.003
Examiner B	0.68	0.52	0.40	0.33	0.04
Relocation*					
Examiner A	0.83	0.55	0.35	0.61	<0.001
Examiner B	0.73	0.63	0.25	0.27	0.08
Surprise					
Examiner A	0.75	0.53	0.38	0.47	0.003
Examiner B	0.68	0.52	0.40	0.33	0.04
Load and Shift					
Examiner A	0.75	0.65	0.20	0.29	0.03
Examiner B	0.70	0.67	0.18	0.10	0.33
Sulcus sign*					
Examiner A	0.63	0.61	0.28	0.05	0.52
Examiner B	0.63	0.54	0.38	0.19	0.19
Gagey					
Examiner A	0.70	0.60	0.28	0.26	0.10
Examiner B	0.70	0.58	0.30	0.29	0.08

233 Discussion

234 The inter-tester reliability across the selected six clinical shoulder instability and laxity
235 tests ranged from fair to substantial. Use of PABAK calculations improved inter-tester
236 reliability to substantial and almost perfect across most tests, except for the relocation
237 and sulcus sign tests. The tests most often used to characterize self-reported shoulder
238 problems (mutual dependency) were apprehension, relocation and surprise tests.
239
240 The inter-tester reliability for the apprehension, relocation and surprise was higher
241 than, or equivocal, to previously reported results of these tests using the same
242 diagnostic procedures (apprehension and/or pain).²³ Specifically for the apprehension
243 and surprise test, the present k values were somewhat higher than previously reported
244 (0.65 vs. 0.44-0.45). The reason for this may be that the current study included both
245 shoulder cases and shoulder healthy individuals as opposed to only including
246 symptomatic subjects.²³ This may have increased subject variation, known to affect
247 reliability positively. Also, PABAK calculations did not affect the overall reliability of the
248 apprehension and surprise tests, probably due to an optimal prevalence index of
249 positive and negative tests (close to 0.50). For the relocation test, the existing inter-
250 tester reliability was almost similar to previously reported (k 0.39 vs. 0.44),²³ however,
251 fairly lower. Apparently, the primary reason for the current poor reliability in
252 relocation was presence of systematic bias between testers, as indicated by the actual
253 raw data (contingency tables) and the statistical significant inter-examiner difference.
254 Likewise, systematic bias between testers was also found for the sulcus sign test in the
255 present study. Hypothetically, this may be explained by inter-tester variability in the

Inter-tester reliability of clinical shoulder instability and laxity tests

1
2
3
4 256 force produced to relocate, respectively inferiorly translate the glenohumeral head
5
6 257 during the relocation and sulcus sign tests in the current study. This is, however, only
7
8
9 258 speculative and further studies are needed to standardize these tests.
10
11 259 Reliability for the present sulcus sign test was slightly lower than previously reported (k
12
13 260 0.39 vs. >0.50).^{22,23} The discrepancy in reliability observed may be due to the use of
14
15 261 different test positions with participants in the current study sitting upright²⁹ as
16
17 262 opposed to a previous lying test position.²² Although, using PABAK it applies for both
18
19 263 the relocation and sulcus sign test that overall reliability was not affected much, due to
20
21 264 the presence of systematic bias.
22
23 265 For the load-and-shift test, reliability was relatively low (including wide CI). This may be
24
25 266 due to the current poor prevalence index. However, the present dichotomous rating of
26
27 267 the load-and-shift test (meaning that only individuals that could either subluxate or
28
29 268 dislocate the shoulder during testing was deemed positive) may have influenced the
30
31 269 prevalence of positive tests largely. Therefore, using PABAK, reliability of the load-and-
32
33 270 shift test improved considerably (from moderate to almost perfect). Nevertheless,
34
35 271 different statistics (kappa vs. Intra-class-Correlation Coefficients), different scoring
36
37 272 systems (dichotomous rating (positive yes/no) vs. four point grading scale (0-3)²³ and
38
39 273 inclusion of shoulder asymptomatic athletes only²² make comparison across studies
40
41 274 difficult.
42
43 275 Finally, reliability of the Gagey test was substantial and PABAK did not affect reliability
44
45 276 much due to a nearly optimal prevalence and low bias between testers. Unfortunately,
46
47 277 there is no other study to compare with.
48
49
50
51
52
53
54
55
56
57
58
59
60

Inter-tester reliability of clinical shoulder instability and laxity tests

1
2
3
4 278 Although the current study was designed to investigate reliability, and not diagnostic
5
6 279 accuracy, the mutual dependency between the individual tests and self-reported
7
8
9 280 shoulder problems was analysed. It revealed that the tests most often used to
10
11 281 characterize those with and without self-reported shoulder problems (mutual
12
13 282 dependency) proved to be the apprehension, relocation and surprise tests. This may
14
15 283 indicate a relationship between these tests, which may come as no surprise, since
16
17 284 these tests are a continuum of the apprehension test and, thus, closely related.⁹
18
19 285 Nevertheless, for clinicians it is of interest to specify the clinical characteristics of
20
21 286 patients with self-reported shoulder problems. Thus, the current prevalence of
22
23 287 positive tests may mirror these characteristics of the included patients and should be
24
25 288 taken into consideration in the management of such musculoskeletal conditions. It
26
27 289 could be suggested to develop and test the clinimetric properties of a more
28
29 290 comprehensive test battery for evaluating self-reported shoulder problems. No prior
30
31 291 studies were found addressing mutual dependency of the current tests for shoulder
32
33 292 instability and laxity, which hampers comparison.
34
35 293 The present study has several limitations. First, despite the use of a standardized test
36
37 294 protocol, including training and an overall agreement phase, two of the tests
38
39 295 (relocation and sulcus sign tests) failed to reach recommendable reliability values.
40
41 296 Further standardization in both performance and interpretation are therefore needed.
42
43 297 Secondly, no valid gold standard for classifying shoulder instability/laxity was used. To
44
45 298 compensate for this, self-reported confirmation of shoulder-related problems was
46
47 299 applied, but this was not reflected in the current WOSI scores, which were relatively
48
49
50
51
52
53
54
55
56
57
58
59
60

Inter-tester reliability of clinical shoulder instability and laxity tests

1
2
3
4 300 low. Lack of a more objective gold standard may have decreased diagnostic accuracy,
5
6 301 however, not reliability, which was the primary objective of the present study-
7
8 302 Thirdly, although the recommended prevalence of 0.50 in reliability studies²⁵ in all six
9
10 303 tests was not accomplished, use of PABAK calculations was able to represent a valid
11
12 304 alternative for the original kappa values. Nevertheless, future studies should use
13
14 305 inclusion criteria of more established shoulder instability and laxity conditions, and, if
15
16 306 possible, verified by objective criteria as surrogate for a gold standard of shoulder
17
18 307 instability and laxity. This may optimize prevalence as well as diagnostic accuracy in
19
20 308 studies where this is a further aim.
21
22 309 The strengths of the study are the use of standardized procedures (including blinding
23
24 310 to patient status and the use of a three-phased protocol for conducting reliability
25
26 311 studies). Also, presentation of raw data, using contingency tables, along with kappa
27
28 312 and PABAK values, increases data transparency and improves interpretation of the
29
30 313 reliability study.
31
32
33
34
35
36
37

314 **Conclusions**

315 This study showed acceptable inter-tester reliability for four of six clinical shoulder
316 instability and laxity tests in relatively sports active individuals with and without self-
317 reported shoulder disability. However, relocation and sulcus sign tests need further
318 standardisation before being recommended for use in clinical practice. Based on the
319 frequency and mutual dependency of the current tests, especially apprehension and
320 surprise tests seem important in the characterisation of self-reported shoulder
321
322
323
324
325
326
327
328
329
330

Inter-tester reliability of clinical shoulder instability and laxity tests

1
2
3
4 321 problems. Future research on the validity of tests for shoulder instability and laxity is

5
6 322 needed.

7
8 323

9
10 324

11
12 325

13
14 326

15
16 327

17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

Inter-tester reliability of clinical shoulder instability and laxity tests

1
2
3
4 328 **Author statement:** HE, KGI, CML and BJK conceived and designed the study. HE and
5
6 329 BHK recruited study participants and collected data. HE performed the statistical
7
8 330 analysis. HE, KGI, CML, BJK interpreted the results. HE drafted the manuscript with
9
10 331 KGI, CML, BHK and BJK contributing to the manuscript. All authors have read and
11
12 332 approved the final manuscript. HE is the guarantor.
13
14
15
16

17 334 **Funding:** This work was supported by Region of Southern Denmark's Research fund
18
19 335 and The Danish Rheumatism Association.
20
21
22

23 336
24 337 **Competing interest:** None declared
25
26
27

28 339 **Data sharing statement:** No additional data are available
29
30
31 340
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

References

1. Luime, J. J., B. W. Koes, I. J. Hendriksen, A. Burdorf, A. P. Verhagen, H. S. Miedema, and J. A. Verhaar. 2004. 'Prevalence and incidence of shoulder pain in the general population; a systematic review', *Scandinavian Journal of Rheumatology*, 33: 73-81.
2. Farrar, N. G., J. J. Malal, J. Fischer, and M. Waseem. 2013. 'An overview of shoulder instability and its management', *Open Orthopaedics Journal*, 7: 338-46.
3. Goss, T. P. 1988. 'Anterior glenohumeral instability', *Orthopedics*, 11: 87-95.
4. Cameron, K. L., M. L. Duffey, T. M. DeBerardino, P. D. Stoneman, C. J. Jones, and B. D. Owens. 2010. 'Association of generalized joint hypermobility with a history of glenohumeral joint instability', *J Athl Train*, 45: 253-8.
5. Chahal, J., J. Leiter, M. D. McKee, and D. B. Whelan. 2010. 'Generalized ligamentous laxity as a predisposing factor for primary traumatic anterior shoulder dislocation', *J Shoulder Elbow Surg*, 19: 1238-42.
6. Dodson, C. C., and F. A. Cordasco. 2008. 'Anterior glenohumeral joint dislocations', *Orthop Clin North Am*, 39: 507-18, vii.
7. Jaggi, A., and S. Lambert. 2010. 'Rehabilitation for shoulder instability', *Br J Sports Med*, 44: 333-40.
8. Wilk, K. E., L. C. Macrina, and M. M. Reinold. 2006. 'Non-operative rehabilitation for traumatic and atraumatic glenohumeral instability', *N Am J Sports Phys Ther*, 1: 16-31.
9. Hegedus, E. J., A. P. Goode, C. E. Cook, L. Michener, C. A. Myer, D. M. Myer, and A. A. Wright. 2012. 'Which physical examination tests provide clinicians with the most value when examining the shoulder? Update of a systematic review with meta-analysis of individual tests', *Br J Sports Med*, 46: 964-78.
10. Ann Ganestam, Mikkel L. Attrup, Per Hølmich and Kristoffer W Barfoed. 2015. 'Evaluation of Clinical Practice of Shoulder Examination among Ten Experienced Shoulder Surgeons', *Journal of Orthopaedic Research and Physiotherapy*, 1.
11. Luime, J. J., A. P. Verhagen, H. S. Miedema, J. I. Kuiper, A. Burdorf, J. A. Verhaar, and B. W. Koes. 2004. 'Does this patient have an instability of the shoulder or a labrum lesion?', *JAMA*, 292: 1989-99.
12. Tzannes, A., and G. A. Murrell. 2002. 'Clinical examination of the unstable shoulder', *Sports Medicine*, 32: 447-57.
13. Gagey, O. J., and N. Gagey. 2001. 'The hyperabduction test', *J Bone Joint Surg Br*, 83: 69-74.

Inter-tester reliability of clinical shoulder instability and laxity tests

- 1
2
3
4 388 14. Lo, I. K., B. Nonweiler, M. Woolfrey, R. Litchfield, and A. Kirkley. 2004. 'An
5 389 evaluation of the apprehension, relocation, and surprise tests for anterior shoulder
6 390 instability', *Am J Sports Med*, 32: 301-7.
7 391
8 392 15. Silliman, J. F., and R. J. Hawkins. 1993. 'Classification and physical diagnosis of
9 393 instability of the shoulder', *Clin Orthop Relat Res*: 7-19.
10 394
11 395 16. Cavallo, R. J., and K. P. Speer. 1998. 'Shoulder instability and impingement in
12 396 throwing athletes', *Med Sci Sports Exerc*, 30: S18-25.
13 397
14 398 17. Matsen, F. A., 3rd, and J. D. Zuckerman. 1983. 'Anterior glenohumeral
15 399 instability', *Clin Sports Med*, 2: 319-38.
16 400
17 401 18. Robinson, C. M., J. Howes, H. Murdoch, E. Will, and C. Graham. 2006. 'Functional
18 402 outcome and risk of recurrent instability after primary traumatic anterior shoulder
19 403 dislocation in young patients', *J Bone Joint Surg Am*, 88: 2326-36.
20 404
21 405 19. van Kampen, D. A., T. van den Berg, H. J. van der Woude, R. M. Castelein, C. B.
22 406 Terwee, and W. J. Willems. 2013. 'Diagnostic value of patient characteristics,
23 407 history, and six clinical tests for traumatic anterior shoulder instability', *J Shoulder
24 408 Elbow Surg*, 22: 1310-9.
25 409
26 410 20. Warby, S. A., T. Pizzari, J. J. Ford, A. J. Hahne, and L. Watson. 2014. 'The effect of
27 411 exercise-based management for multidirectional instability of the glenohumeral
28 412 joint: a systematic review', *J Shoulder Elbow Surg*, 23: 128-42.
29 413
30 414 21. Gibson, K., A. Growse, L. Korda, E. Wray, and J. C. MacDermid. 2004. 'The
31 415 effectiveness of rehabilitation for nonoperative management of shoulder
32 416 instability: a systematic review', *J Hand Ther*, 17: 229-42.
33 417
34 418 22. Levy, A. S., S. Lintner, K. Kenter, and K. P. Speer. 1999. 'Intra- and interobserver
35 419 reproducibility of the shoulder laxity examination', *Am J Sports Med*, 27: 460-3.
36 420
37 421 23. Tzannes, A., A. Paxinos, M. Callanan, and G. A. Murrell. 2004. 'An assessment of
38 422 the interexaminer reliability of tests for shoulder instability', *J Shoulder Elbow
39 423 Surg*, 13: 18-23.
40 424
41 425 24. Kottner, J., L. Audige, S. Brorson, A. Donner, B. J. Gajewski, A. Hrobjartsson, C.
42 426 Roberts, M. Shoukri, and D. L. Streiner. 2011. 'Guidelines for Reporting Reliability
43 427 and Agreement Studies (GRRAS) were proposed', *Int J Nurs Stud*, 48: 661-71.
44 428
45 429 25. Patijn J, Remvig L. . 2007. "Reproducibility And Validity 2007-Protocol Formats
46 430 For Diagnostics Procedures In Manual/Musculoskeletal Medicine." In.
47 431
48 432 26. Downie, W. W., P. A. Leatham, V. M. Rhind, V. Wright, J. A. Branco, and J. A.
49 433 Anderson. 1978. 'Studies with pain rating scales', *Annals of the Rheumatic Diseases*,
50 434 37: 378-81.
51
52
53
54
55
56
57
58
59
60

Inter-tester reliability of clinical shoulder instability and laxity tests

- 1
2
3
4 435
5 436 27. Kirkley, A., S. Griffin, H. McLintock, and L. Ng. 1998. 'The development and
6 437 evaluation of a disease-specific quality of life measurement tool for shoulder
7 438 instability. The Western Ontario Shoulder Instability Index (WOSI)', *Am J Sports*
8 439 *Med*, 26: 764-72.
9 440
10 441 28. 'World Medical Association Declaration of Helsinki: ethical principles for
11 442 medical research involving human subjects'. 2000. *JAMA*, 284: 3043-5.
12 443
13 444 29. Bahk, M., E. Keyurapan, A. Tasaki, E. L. Sauers, and E. G. McFarland. 2007.
14 445 'Laxity testing of the shoulder: a review', *Am J Sports Med*, 35: 131-44.
15 446
16 447 30. Sim, J., and C. C. Wright. 2005. 'The kappa statistic in reliability studies: use,
17 448 interpretation, and sample size requirements', *Phys Ther*, 85: 257-68.
18 449
19 450 31. Byrt, T., J. Bishop, and J. B. Carlin. 1993. 'Bias, prevalence and kappa', *J Clin*
20 451 *Epidemiol*, 46: 423-9.
21 452
22 453 32. Hoehler, F. K. 2000. 'Bias and prevalence effects on kappa viewed in terms of
23 454 sensitivity and specificity', *J Clin Epidemiol*, 53: 499-503.
24 455
25 456 33. Landis, J. R., and G. G. Koch. 1977. 'The measurement of observer agreement for
26 457 categorical data', *Biometrics*, 33: 159-74.
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

BMJ Open

Inter-tester reliability of clinical shoulder instability and laxity tests in subjects with and without self-reported shoulder problems

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2017-018472.R1
Article Type:	Research
Date Submitted by the Author:	06-Oct-2017
Complete List of Authors:	Eshoj, Henrik; Syddansk Universitet Det Sundhedsvidenskabelige Fakultet, Department of Sports Science and Clinical Biomechanics; Odense Universitetshospital, Quality of Life Research Center, Department of Haematology Ingwersen, Kim Gordon; Hospital Lillebaelt - Vejle Hospital, Vejle, Denmark, Research Unit at Department of Physical- and Occupational Therapy,; Syddansk Universitet Det Sundhedsvidenskabelige Fakultet, Department of Sports Science and Clinical Biomechanics Larsen, Camilla; Research Unit for Musculoskeletal Function and Physiotherapy, Institute of Sports Science and Clinical Biomechanics Kjær, Birgitte; Bispebjerg and Frederiksberg University Hospital, Copenhagen, Denmark , Research Unit at Department of Physical- and Occupational Therapy, ; Syddansk Universitet Det Sundhedsvidenskabelige Fakultet, Department of Sports Science and Clinical Biomechanics Juul-Kristensen, Birgit; Research Unit for Musculoskeletal Function and Physiotherapy, Institute of Sports Science and Clinical Biomechanics
Primary Subject Heading:	Evidence based practice
Secondary Subject Heading:	Sports and exercise medicine
Keywords:	Reliability, Shoulder < ORTHOPAEDIC & TRAUMA SURGERY, Instability, Laxity, Clinical test

SCHOLARONE™
Manuscripts

1
2
3
4 1 **Inter-tester reliability of clinical shoulder instability and laxity tests in subjects**
5
6
7 2 **with and without self-reported shoulder problems**
8
9

10 3

11 4 Henrik Eshoj, PhD ^a Kim Gordon Ingwersen, PhD ^{b,c}, Camilla Marie Larsen, PhD ^{b,d} Birgitte Hougs

12 5 Kjær, Msc ^{b,e}, Birgit Juul-Kristensen, PhD ^{b,f}
13
14
15
16
17 6

18
19 7 **Affiliations**

20 8 ^a Quality of Life Research Center, Department of Haematology, Odense University Hospital,
21 9 Odense, Denmark

22 10 ^b Department of Sports Science and Clinical Biomechanics, University of Southern Denmark,
23 11 Odense, Denmark

24 12 ^c Research Unit at Department of Physical- and Occupational Therapy, Hospital Lillebaelt - Vejle
25 13 Hospital, Vejle, Denmark

26 14 ^d Health Sciences Research Centre, University College Lillebaelt, Odense, Denmark

27 15 ^e Research Unit at Department of Physical- and Occupational Therapy, Bispebjerg and
28 16 Frederiksberg University Hospital, Copenhagen, Denmark

29 17 ^f Institute of Occupational Therapy, Physiotherapy and Radiography, Department of Health
30 18 Sciences, Bergen University College, Norway
31
32
33
34
35
36
37
38 19

39 20 **Correspondence:**

40 21 Birgit Juul-Kristensen

41 22 Department of Sports Science and Clinical Biomechanics

42 23 University of Southern Denmark, Campusvej 55, 5230 Odense M, DK

43 24 Tel.: +45 2942 3335

44 25 Email: bjjuul-kristensen@health.sdu.dk
45
46
47
48
49
50
51
52 26

53 27 **Word counts: 3001**
54
55
56
57 28
58
59
60

1
2
3
4 29 **Abstract**

5
6
7 30 **Objective:** Firstly, to investigate the inter-tester reliability of clinical shoulder instability and laxity
8
9 31 tests, and secondly, to describe the mutual dependency of each test evaluated by each tester for
10
11 32 identifying self-reported shoulder instability and laxity.
12
13

14 33
15
16 34 **Method:** A standardized protocol for conducting reliability studies was used to test the inter-
17
18 35 tester reliability of the six clinical shoulder instability and laxity tests; apprehension, relocation,
19
20
21 36 surprise, load-and-shift, sulcus sign and Gagey. Cohens kappa (k) with 95% confidence intervals
22
23 37 (CI) besides Prevalence-Adjusted-Bias-Adjusted-Kappa (PABAK), accounting for insufficient
24
25 38 prevalence and bias, were computed to establish the inter-tester reliability and mutual
26
27 39 dependency.
28
29

30 40
31
32 41 **Results:** Forty individuals (13 with self-reported shoulder instability and laxity related shoulder
33
34 42 problems and 27 normal shoulders individuals) aged 18-60 were included. Fair (relocation),
35
36 43 moderate (load-and-shift, sulcus sign) and substantial (apprehension, surprise, Gagey) inter-tester
37
38 44 reliability were observed across tests (k 0.39-0.73; 95% CI: 0.00-1.00). PABAK improved reliability
39
40 45 across tests, resulting in substantial to almost perfect inter-tester reliability for the apprehension,
41
42 46 surprise, load-and-shift and Gagey tests (k 0.65-0.90). Mutual dependencies between each test
43
44 47 and self-reported shoulder problem showed apprehension, relocation and surprise to be the most
45
46 48 often used tests to characterize self-reported shoulder instability conditions.
47
48
49
50

51 50 **Conclusions:** Four tests (apprehension, surprise, load-and-shift and Gagey) out of six, were
52
53
54 51 considered inter-tester reliable for clinical use, while relocation and sulcus sign tests need further
55
56
57
58
59
60

1
2
3
4 52 standardization before acceptable evidence. Furthermore, the validity of the tests for instability
5
6 53 and laxity needs to be studied.
7
8

9 54

10 11 55 Article summary section

12 56

13 57 Strengths and limitations of this study

14 58

- 15 59
- 16 59 • The strength of the study is the use of a three-phased standardized study protocol
 - 17
18 60 • Presentation of raw findings increases transparency and interpretation of study findings
 - 19 61 • No valid gold standard for including shoulder instability and laxity subjects was used
 - 20
21 61 • No valid gold standard for including shoulder instability and laxity subjects was used
 - 22
23 62 • A 50/50 prevalence of positive and negative tests for all six tests was not accomplished
 - 24
25

26 63

27 28 64 Keywords

29
30 65 Reliability, Shoulder instability, laxity, clinical tests

31
32 66

33 34 67 Acknowledgements

35
36 68 We would like to thank Physiotherapists Rasmus Fitzner, Pernille Madsen and Jacob Hansen from
37
38 69 Metropolitan University College, Copenhagen, Denmark for recruitment and testing of study
39
40 69 Metropolitan University College, Copenhagen, Denmark for recruitment and testing of study
41
42 70 participants. Furthermore, a special thanks to Bispebjerg Frederiksberg University Hospital,
43
44 71 Copenhagen, Denmark for providing facilities for data collection.
45
46

47
48 72

49
50 73

51
52

53

54

55

56

57

58

59

60

74 Introduction

75 Shoulder complaints, affecting shoulder-related quality of life (QoL), are frequent and may be
76 caused by shoulder instability and/or laxity¹ due to traumatic or non-traumatic injuries to the
77 shoulder joint.² The traumatic shoulder instability is mainly prompted by a high impact injury
78 during sports participation, resulting in a shoulder dislocation, predominantly in anterior
79 direction.³ The non-traumatic shoulder instability is usually related to repetitive overhead
80 activities and/or patients with generalised joint hypermobility or glenohumeral hyperlaxity, often
81 referred to as multidirectional shoulder instability.^{2,4,5}

82 Irrespectively of aetiology, shoulder instability and laxity is often accompanied by a variety of
83 symptoms including shoulder discomfort, pain besides glenohumeral subluxations and/or
84 repeated dislocations.^{6,7,8} Clinically, shoulder instability and laxity, are diagnosed and verified by a
85 group of shoulder pain and instability provoking/relief tests, supplemented by shoulder laxity
86 tests.^{9,10} The former tests usually include the anterior shoulder instability and laxity tests;
87 apprehension, relocation and surprise, and the laxity tests consisting of the load-and-shift, sulcus
88 sign and Gagey tests.^{11,12,13} An ongoing discussion is the use of pain as diagnostic criterion in
89 diagnosing anterior shoulder instability with the clinical tests apprehension, relocation and
90 surprise.^{14,15,16} In one way, it may be a confounding factor, since pain has shown to be less
91 predictive and reliable as a diagnostic criterion.¹⁴ On the contrary though, others have suggested
92 that unrecognized and underlying glenohumeral instability may lead to repetitive microtrauma
93 and painful shoulder conditions,^{15,16} justifying pain as diagnostic criterion when testing for anterior
94 shoulder instability.

95 Nonetheless, symptoms may become chronic, and lead to reduced work and sports
96 capability,^{17,18,19} and with exercise-based management as the most often recommended first-

Inter-tester reliability of clinical shoulder instability and laxity tests

1
2
3
4 97 choice treatment.^{20,21} Hence, early diagnosis using reliable and accurate clinical tests to guide
5
6 98 focused treatment is essential. Few studies though, have investigated the reliability of clinical
7
8
9 99 shoulder instability and laxity tests showing large variations in reliability and with limited
10
11 100 methodological quality, hampering interpretation and comparison with other studies.^{14,22,23}
12
13
14 101

15
16 102 Therefore, the objective of this study was to investigate the inter-tester reliability of commonly
17
18 103 used clinical shoulder instability and laxity tests and secondly to describe the mutual dependency
19
20 104 for each test evaluated by each tester, in a group of sports-active individuals with and without
21
22
23 105 self-reported shoulder problems.
24
25 106

107 **Materials and methods**

108 **Study design**

109 An inter-tester reliability study was conducted involving two physiotherapists as inter-tester
110 examiners. A third physiotherapist (study coordinator), not involved in the actual inter-tester
111 reliability study (test phase), managed all practical aspects during the study period. The Guidelines
112 for Reporting Reliability and Agreement Studies (GRASS), a consensus document on how to report
113 reliability and agreement studies, were followed.²⁴ A standardized protocol for reliability studies,
114 consisting of three phases: preparation and training of clinical tests, overall agreement, and test
115 phase (the actual reliability study) were applied.²⁵ Two early career physiotherapists with six
116 months clinical experience were involved in the inter-tester reliability study. A test-protocol
117 describing each clinical test was developed and subsequently used by the two testers to practice
118 all tests in order to reach uniformity and mutual agreement in performing and interpreting each
119 test. In the overall agreement phase, the two testers examined 19 individuals (eight affected
120 shoulders and 11 normal shoulders). The two testers were mutually blinded to the health status of
121 the individuals (affected shoulders vs. normal shoulders) and also to each other's test results.
122 Before proceeding to the final study phase, the two testers needed an overall agreement of at
123 least 80% based on findings from the six clinical shoulder tests.²⁵ In the actual inter-tester
124 reliability test phase, the two testers examined a new group of individuals with affected,
125 respectively normal shoulders with the six clinical shoulder tests. The procedure was the same as
126 in the agreement phase, meaning that testers were blinded to the health status of the individuals
127 and each other's test results.

128

129

1
2
3
4 130 **Study subjects**

5
6 131 A sample size of at least 40 individuals was targeted based on recommendations for performing
7
8 132 clinical reliability studies.²⁵ Sixty-five individuals (women and men (aged 18-60 years)) were
9
10 133 recruited and screened for eligibility from Metropolitan University College, Copenhagen, and
11
12 134 Bispebjerg Frederiksberg University Hospital, Copenhagen resulting in an included number of 13
13
14 135 individuals with instability and/or laxity related shoulder problems (hereinafter referred to as
15
16 136 shoulder affected) vs. 27 normal shoulder individuals, respectively.
17
18 137 Shoulder affected individuals answering yes to at least one of two questions (*'Do you have a sense*
19
20 138 *of shoulder instability?'* and *'Have you ever had a shoulder injury?'*) were eligible for a clinical
21
22 139 shoulder examination performed by the study coordinator. The shoulder affected individuals were
23
24 140 then included if they present with at least one positive clinical shoulder test out of the following;
25
26 141 apprehension, relocation, surprise, load-and-shift, sulcus sign or Gagey. Individuals with normal
27
28 142 shoulders were recruited through public advertisements followed by a telephone interview and
29
30 143 included if they present with no self-reported shoulder pathology or complaints. In general, any
31
32 144 individuals with prior shoulder surgery were excluded. In the actual test phase, individuals
33
34 145 completed a short questionnaire with basic demographic details (age, gender, weight, height), in
35
36 146 addition to the following: pain level during rest and activity (numeric pain rating scale, NPRS),²⁶
37
38 147 shoulder injury ever (yes/no), subjective shoulder instability (yes/no) and sports-related activity
39
40 148 (hours/week). Further, all individuals completed the patient-reported Western Ontario Shoulder
41
42 149 Instability (WOSI) questionnaire designed to measure shoulder function and quality of life in
43
44 150 patients with shoulder instability and laxity symptoms.²⁷ The time period between each test phase
45
46 151 was approximately 2 weeks, and new subjects were included for each phase. Only the study phase
47
48 152 is reported in the current manuscript. The study was exempted for notification to the Danish
49
50
51
52
53
54
55
56
57
58
59
60

Inter-tester reliability of clinical shoulder instability and laxity tests

1
2
3
4 153 Health Research Study Board due to the non-invasive and non-treating study design. However,
5
6 154 oral and written consent was provided from all individuals and, ethical guidelines were followed
7
8
9 155 according to the Helsinki declaration.²⁸
10

11 156

12
13 157 **Clinical tests**

14
15
16 158 The clinical shoulder tests consisted of three shoulder joint-provoking tests for anterior shoulder
17
18 159 instability (apprehension, relocation and surprise) besides three shoulder laxity tests (load and
19
20 160 shift, sulcus sign and Gagey) (Table 1).^{11,13,14,22,23,29}
21

22
23 161 The apprehension test (Table 1, Figure 1) was positive if glenohumeral apprehension and/or pain
24
25 162 were evoked during testing whereas relief of symptoms with the relocation test (Table 1, Figure 2)
26
27 163 was regarded as a positive test. As for the apprehension, the surprise test (Table 1, Figure 3) was
28
29 164 positive if glenohumeral apprehension and/or pain were evoked during testing. The load and shift
30
31 165 test (Table 1, Figure 4 & 5) was rated on a four-point scale ranging from 0-3 (best to worst; 0= little
32
33 166 glenohumeral movement; 3=humeral head moves beyond the glenoid rim and remains
34
35 167 dislocated).¹² Also, to enhance mutual agreement between testers when performing the load-and-
36
37 168 shift test, only the direction (anterior vs. posterior) with most glenohumeral head translation was
38
39 169 scored. Sulcus sign (Table 1, Figure 6) was objectively measured in centimeter (continuous scale)
40
41 170 by use of a small ruler according to previously used grading scales as follows; I (<1 cm translation),
42
43 171 II (1 to 2.0 cm translation) or III (>2.0 cm translation).²⁹ Finally, Gagey test (Table 1, Figure 7) was
44
45 172 rated as positive with passive abduction above 105 degrees.¹³
46
47
48

49 173

50
51
52 174 **Statistics**

53
54
55
56 175 Demographics and descriptive data were tested for normality by visual inspection of histograms
57
58
59
60

Inter-tester reliability of clinical shoulder instability and laxity tests

1
2
3
4 176 and Shapiro-Wilk's test. Group differences (affected shoulders vs. normal shoulders) were tested
5
6 177 by Fisher's Exact test for categorical variables, whereas student's t-test and Mann-Whitney U-test
7
8
9 178 was used for parametric, respectively non-parametric distributed data.
10
11 179 Apprehension, relocation, surprise and Gagey tests were dichotomous variables whereas the load
12
13 180 and shift and sulcus sign tests were dichotomized to also allow for nominal statistics. Thus, load-
14
15
16 181 and-shift was rated positive when scored 2 or 3, while for sulcus sign a positive rating was equal to
17
18 182 measurements exceeding one centimeter.²⁹ For transparency, data from each test is presented by
19
20
21 183 2 x 2 contingency tables besides the use of McNemar's test for significant between-tester
22
23 184 differences. Furthermore, observed and expected agreements are presented along with
24
25 185 prevalence and bias³⁰ indexes. Reliability was evaluated with the use of Cohen's kappa (k)
26
27
28 186 coefficients including 95% confidence intervals (CI).²⁵ Also, since kappa is sensitive to imbalances
29
30 187 in prevalence and bias (e.g. if a 50/50 distribution of positive and negative tests cannot be
31
32 188 accomplished) the use of PABAK calculation is a valid supplement to the original kappa values.^{30,31}
33
34
35 189 By definition, PABAK reflects the ideal situation, thereby accounting for variation of prevalence
36
37 190 and bias between testers (as presented in the "real" world).³² PABAK calculation is performed by
38
39 191 adjusting for high or low prevalence by computing the average of cells a and d in a cross table,
40
41 192 substituting this value for the actual values in those cells. Similarly, an adjustment for bias is
42
43
44 193 achieved by substituting the mean of cells b and c for those actual cell values.³⁰ Finally, the
45
46 194 relationship for each tester between the individual tests and the classification (mutual
47
48
49 195 dependency) by self-reported shoulder problems was tested by Cohen's kappa (k) coefficients and
50
51 196 the characterization of the groups was tested with Fischer's exact tests.
52
53 197 The classification system proposed by Landis and Koch was used to interpret reliability as follows:
54
55
56 198 0.00-0.20 (Slight); 0.21-0.40 (Fair); 0.41-0.60 (Moderate); 0.61-0.80 (Substantial) and 0.81-1.00
57
58
59
60

Inter-tester reliability of clinical shoulder instability and laxity tests

1
2
3
4 199 (Almost perfect).³³
5

6 200 Statistical Package for the Social Sciences (SPSS inc., Chicago, IL, USA), version 22, was used for all
7

8
9 201 statistical analyses, with p-value of <0.05 interpreted as significant.
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

Inter-tester reliability of clinical shoulder instability and laxity tests

Table 1. Performance and evaluation of the clinical shoulder instability and laxity tests**Verbal introduction:**

1. I am going to perform six clinical shoulder tests on you
2. I will ask if you experience any symptoms (apprehension and/or pain) during the three first tests.
3. I will guide you through each test

Clinical tests	Description	Placing of hands, etc.	Evaluation
Figure 1	Individuals placed supine with the shoulder being tested close to the edge of the examination table.	One hand around the wrist of the individual with the other hand gently placed in front of the shoulder.	Subjective or objective presence of apprehension and/or pain?
	Shoulder positioned in 90° of abduction, elbow flexed to 90°.	Elbow supported at the examiners thigh.	Rated as either positive or negative.
	Examiner moves the shoulder into maximal external rotation.		(Nominal, dichotome data)
Figure 2	From the end position of the apprehension test the humeral head is gently forced posteriorly	Examiners fifth finger placed close to the lateral part of the acromion with the wrist positioned anteriorly at the humeral head.	Relief of apprehension and/or pain? Rated as either positive or negative. (Nominal, dichotome data)
	From end position of the relocation test the posteriorly directed force at the humeral head is quickly removed.	Removal of examiners wrist from the anterior part of the shoulder.	Subjective or objective reproduction of apprehension and/or pain? Rated as either positive or

Inter-tester reliability of clinical shoulder instability and laxity tests

			negative.
			(Nominal, dichotome data)
Figure 4	Individual placed supine with scapula resting at the examination table. Humeral head is loaded gently into the glenoid through axial pressure at the elbow.	Examiners one hand placed at the olecranon with the individual's hand positioned between the examiners torso and elbow.	Humeral head movement evaluated by the use of a four-level laxity scale. 0 = little to almost no movement 1 = humeral head moves up onto the glenoid
Figure 5	<i>Anterior load-and-shift test:</i> Shoulder positioned in the scapular plane in 90° of abduction with elbow flexed. Humeral head gently shifted in anterior direction. <i>Posterior load-and-shift test:</i> Shoulder positioned in the scapular plane in 20° of abduction with elbow flexed. Humeral head gently shifted in posterior direction.	<i>Anterior load-and-shift:</i> Examiners hand placed on top of the shoulder with the fingers on the backside of the glenohumeral head to move it anteriorly. <i>Posterior load-and-shift:</i> Examiners wrist placed at the anterior part of the humeral head to move it posteriorly.	2 = humeral head moves beyond the glenoid, but relocates spontaneously once pressure is released 3 = humeral head moves beyond the glenoid and remains dislocated Rated as positive when scored 2 or 3. (Nominal, dichotomous data)
Figure 6	Individuals sitting upright. Shoulder in neutral position (0 degree rotation).	One hand placed above the epicondyles of humerus.	Rated as positive with measurements exceeding 1 centimeter.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

Inter-tester reliability of clinical shoulder instability and laxity tests

	Examiner pulls the distal part of the humerus in a caudal direction.	Other hand is used to evaluate the subacromial distance with a ruler.	(Nominal, dichotome data)
	Distance from the top of the humeral head and the acromion is evaluated with a ruler.		
Figure 7	Individuals sitting upright. The shoulder girdle is stabilized by examiners forearm preventing the shoulder girdle to elevate while the individuals arm is passively moved into end range in horizontal abduction.	Forearm placed on top of the shoulder girdle with the other hand placed around the elbow joint.	Rated as positive with abduction exceeding 105 degrees. (Nominal, dichotomous data)
	A mirror in front of the individual is used to evaluate the shoulder abduction angle.		

Inter-tester reliability of clinical shoulder instability and laxity tests

202 **Results**

203 Characteristics of the participating individuals are presented in Table 2. Demographics
 204 showed no difference between the individuals with affected shoulders (n=13) and
 205 normal shoulders (n=27). Furthermore, both groups (92 and 74%; p=0.18) were
 206 relatively active with a weekly participation in sports-related activity for more than
 207 four hours per week. However, as expected due to the design, affected shoulders had
 208 significantly higher pain during activity (4.23 vs. 1.44; p=0.02), higher frequency of
 209 shoulder injury ever (62% vs. <1%; p<0.001), higher subjective shoulder instability (69
 210 vs. 11%; p<0.001) and worse total WOSI score (506 vs. 136; p=0.001) (Table 2).

212 **Table 2.** Study phase. Participant characteristics.

	Affected shoulders (n=13)	Normal shoulders (n=27)	P-value
Sex (women/men)	8/5	21/6	0.28
Age (years) mean (SD)	28 (9)	29 (7)	0.72
Weight (kg)* mean (SD)	71.0 (12.8)	74.9 (23.4)	0.59
Height (cm) mean (SD)	174.0 (8.6)	173.4 (7.9)	0.82
Pain, rest (NRS 0-10) mean (SD)	1.08 (1.44)	0.41 (1.15)	0.12
Pain, activity (NRS 0-10) mean (SD)	4.23 (2.92)	1.44 (2.12)	<0.05
Shoulder injury ever, n (%)	8 (62)	1 (4)	<0.001
Subjective shoulder instability, n (%)	9 (69)	3 (11)	<0.001
Sports-related activity (>4 hours/week), n (%)	12 (92)	20 (74)	0.18
WOSI domains, mean (SD)			
- Physical symptoms (0-1000)	225 (165)	60 (78)	<0.05
- Sports, recreation, work (0-400)	103 (93)	24 (47)	<0.05
- Lifestyle (0-400)	58 (57)	13 (21)	<0.05
- Emotions (0-300)	121 (94)	39 (49)	<0.05
WOSI total score (0-2100) mean (SD)	506 (362)	136 (174)	<0.001

SD Standard deviation; kg kilo; * significance level p < 0.05; cm centimeter NRS Numeric Rating Scale

213

214

215

Inter-tester reliability of clinical shoulder instability and laxity tests

216 Prevalence of positive tests was especially low for the load-and-shift test (table 3), and
 217 significant between-tester differences were found for relocation and sulcus sign tests
 218 ($p=0.021$) (not shown in tables).

219

220 **Table 3.** Contingency tables with findings from tester A and B.

Apprehension		A	
		Yes	No
B	Yes	14	4
	No	3	19

Relocation		A	
		Yes	No
B	Yes	6	2
	No	8	24

Surprise		A	
		Yes	No
B	Yes	14	4
	No	3	19

Load-and-shift		A	
		Yes	No
B	Yes	1	0
	No	2	37

Sulcus		A	
		Yes	No
B	Yes	7	1
	No	9	23

Gagey		A	
		Yes	No
B	Yes	8	3
	No	1	28

221

222

223 Reliability varied between k : 0.39-0.73 (95% CI: 0.00-1.00), indicating fair (relocation; k
 224 0.39), moderate (load-and-shift, sulcus sign; k 0.43 and 0.48) and substantial
 225 (apprehension, surprise, Gagey; k 0.65-0.73) reliability (Table 4). The prevalence index
 226 of all six tests ranged from 0.05-0.44, (lowest for load-and-shift, relocation and sulcus;
 227 0.05, 0.28 and 0.30), whereas the bias index ranged from and 0.03-0.20 (highest for
 228 relocation and sulcus). PABAK improved reliability for relocation, load-and-shift, sulcus
 229 sign and Gagey test, now corresponding to moderate (relocation and sulcus sign; k
 230 0.50), substantial (Gagey; k : 0.80) and almost perfect (load and shift; k : 0.90) reliability
 231 (Table 4).

232

233

Inter-tester reliability of clinical shoulder instability and laxity tests

Table 4. Reliability of six clinical shoulder instability and laxity tests.

	Observed agreement	Expected agreement	Prevalence index	Bias index	Kappa (95% CI)	PABAK
Apprehension	0.83	0.51	0.44	0.03	0.65 (0.38; 0.85)	0.65
Relocation*	0.75	0.59	0.28	0.15	0.39 (0.07; 0.68)	0.50
Surprise	0.83	0.51	0.44	0.03	0.65 (0.38; 0.85)	0.65
Load and Shift	0.95	0.90	0.05	0.05	0.48 (0.00; 1.00)	0.90
Sulcus sign*	0.75	0.56	0.30	0.20	0.43 (0.17; 0.72)	0.50
Gagey	0.90	0.62	0.40	0.05	0.73 (0.46; 0.94)	0.80

*Significant inter-tester differences, CI Confidence intervals, PABAK prevalence-and-bias-adjusted kappa

The *k* values for mutual dependency indicate that apprehension, relocation and surprise tests for both examiners were the most frequently used tests for characterizing self-reported shoulder problems (Table 5). This was further confirmed by the significant group difference in the presence of positive tests.

Table 5. Kappa statistics for mutual dependency of the individual tests and self-reported shoulder problems for each tester.

	Observed agreement	Expected agreement	Prevalence index	Kappa	p-value (AS/NS)
Apprehension					
Examiner A	0.75	0.53	0.38	0.47	0.003
Examiner B	0.68	0.52	0.40	0.33	0.04
Relocation*					
Examiner A	0.83	0.55	0.35	0.61	<0.001
Examiner B	0.73	0.63	0.25	0.27	0.08
Surprise					
Examiner A	0.75	0.53	0.38	0.47	0.003
Examiner B	0.68	0.52	0.40	0.33	0.04
Load and Shift					
Examiner A	0.75	0.65	0.20	0.29	0.03
Examiner B	0.70	0.67	0.18	0.10	0.33
Sulcus sign*					
Examiner A	0.63	0.61	0.28	0.05	0.52
Examiner B	0.63	0.54	0.38	0.19	0.19
Gagey					
Examiner A	0.70	0.60	0.28	0.26	0.10
Examiner B	0.70	0.58	0.30	0.29	0.08

AS affected shoulders; NS normal shoulder; *Significant inter-tester differences

243 Discussion

244 The inter-tester reliability across the selected six clinical shoulder instability and laxity
245 tests ranged from fair to substantial. Use of PABAK calculations improved inter-tester
246 reliability to substantial and almost perfect across most tests, except for the relocation
247 and sulcus sign tests. The tests most often used to characterize self-reported shoulder
248 instability and laxity (mutual dependency) were apprehension, relocation and surprise
249 tests.

250
251 The inter-tester reliability for the apprehension, relocation and surprise was higher
252 than, or equivalent, to previously reported results of these tests using the same
253 diagnostic procedures (apprehension and/or pain).²³ Specifically for the apprehension
254 and surprise test, the present *k* values were somewhat higher than previously reported
255 (0.65 vs. 0.44-0.45). The reason for this may be that the current study included both
256 affected and normal shoulder individuals as opposed to only including symptomatic
257 subjects.²³ This may have increased subject variation, known to affect reliability
258 positively. Also, PABAK calculations did not affect the overall reliability of the
259 apprehension and surprise tests, probably due to an optimal prevalence index of
260 positive and negative tests (close to 0.50). For the relocation test, the existing inter-
261 tester reliability was almost similar to previously reported (*k* 0.39 vs. 0.44),²³ however
262 lower. Apparently, the primary reason for the current poor reliability in relocation was
263 presence of systematic bias between testers, as indicated by the actual raw data
264 (contingency tables) and the statistical significant inter-examiner difference. Likewise,
265 systematic bias between testers was also found for the sulcus sign test in the present

Inter-tester reliability of clinical shoulder instability and laxity tests

1
2
3
4 266 study. Hypothetically, this may be explained by inter-tester variability in the force
5
6 267 produced to translate the humeral head in posterior (relocation test) or inferior (sulcus
7
8 268 sign test) direction, in the current study. This is, however, only speculative and further
9
10 269 studies are needed to standardize these tests.

11
12
13 270 Reliability for the present sulcus sign test was slightly lower than previously reported (k
14
15 271 0.39 vs. >0.50).^{22,23} The discrepancy in reliability observed may be due to the use of
16
17 272 different test positions with participants in the current study sitting upright²⁹ as
18
19 273 opposed to a previous lying test position.²² However, due to the presence of
20
21 274 systematic bias in both the relocation and sulcus sign test, PABAK did not affect the
22
23 275 overall reliability much.

24
25
26
27 276 For the load-and-shift test, reliability was relatively low (including wide CI). This may be
28
29 277 due to the current low prevalence index below 50%, which is the optimum prevalence
30
31 278 in reliability studies.²⁵ However, the present dichotomous rating of the load-and-shift
32
33 279 test (meaning that only individuals that could either subluxate or dislocate the
34
35 280 shoulder during testing was deemed positive) may have influenced the prevalence of
36
37 281 positive tests largely. Therefore, using PABAK, reliability of the load-and-shift test
38
39 282 improved considerably (from moderate to almost perfect). Nevertheless, different
40
41 283 statistics (kappa vs. Intra-class-Correlation Coefficients), different scoring systems
42
43 284 (dichotomous rating (positive yes/no) vs. four point grading scale (0-3)²³ and inclusion
44
45 285 of shoulder asymptomatic athletes only²² make comparison across studies difficult.

46
47 286 Finally, reliability of the Gagey test was substantial and PABAK did not affect reliability
48
49 287 much due to a nearly optimal prevalence and low bias between testers. Unfortunately,
50
51 288 there is no other study to compare with.
52
53
54
55
56
57
58
59
60

Inter-tester reliability of clinical shoulder instability and laxity tests

1
2
3
4 289 Although the current study was designed to investigate reliability, and not diagnostic
5
6 290 accuracy, the mutual dependency between the individual tests and self-reported
7
8 291 shoulder problems was analysed. It revealed that the tests most often used to
9
10 292 characterize those with and without self-reported shoulder instability and laxity
11
12 293 (mutual dependency) proved to be the apprehension, relocation and surprise tests.
13
14 294 This may indicate a relationship between these tests, which may come as no surprise,
15
16 295 since these tests are a continuum of the apprehension test and, thus, closely related.⁹
17
18 296 Nevertheless, for clinicians it is of interest to specify the clinical characteristics of
19
20 297 patients with self-reported shoulder problems. Thus, the current prevalence of
21
22 298 positive tests may mirror these characteristics of the included patients and should be
23
24 299 taken into consideration in the management of such musculoskeletal conditions. It is
25
26 300 recommended to develop and test the clinimetric properties of a more comprehensive
27
28 301 test battery for evaluating such self-reported shoulder problems. No prior studies were
29
30 302 found addressing mutual dependency of the current tests for shoulder instability and
31
32 303 laxity, which hampers comparison.
33
34 304 The present study has several limitations. Firstly, the lack of standardized
35
36 305 measurement of the amount of force exerted by the two testers during especially the
37
38 306 relocation and sulcus sign test may have limited the current inter-tester reliability.
39
40 307 Further standardization in both performance and interpretation is therefore needed.
41
42 308 Also, the current study did not randomize the order of the clinical tests. However, we
43
44 309 do not believe this to have biased the reliability of the data, since the same order was
45
46 310 used for both testers.
47
48 311 Secondly, no valid gold standard for classifying shoulder instability and laxity was used.
49
50
51
52
53
54
55
56
57
58
59
60

Inter-tester reliability of clinical shoulder instability and laxity tests

1
2
3
4 312 To compensate for this, self-reported confirmation of shoulder-related problems was
5
6 313 applied, but this was not reflected in the current WOSI scores, which were relatively
7
8 314 low. Lack of a more objective gold standard may have decreased diagnostic accuracy,
9
10 315 however not reliability, which was the primary objective of the present study. Also, in
11
12 316 the group with normal shoulders, one individual reported to have had a previous
13
14 317 shoulder injury and three individuals reported subjective shoulder instability, which
15
16 318 does not comply with the inclusion criteria for being regarded as shoulder healthy in
17
18 319 the current study. At the clinical session, a self-reported questionnaire was completed
19
20 320 regarding demographic data and historical information. Apparently, in the baseline
21
22 321 questionnaire three shoulder healthy individuals answered yes to perceiving instability
23
24 322 in their shoulder and one had had a previous shoulder injury, even though they all had
25
26 323 reported no shoulder trouble during the telephone inclusion interview. However, as
27
28 324 depicted in table 2, WOSI and pain scores in the group with normal shoulders seem
29
30 325 not to be influenced severely by these four individuals. Also, re-calculations of
31
32 326 demographic data and mutual dependency with the revised classification into
33
34 327 affected/normal shoulders did not change the mutual dependency of the most
35
36 328 frequently used tests for classification into affected/normal shoulders, and neither was
37
38 329 kappa and demographics affected (data not shown).
39
40 330 Thirdly, due to a relative short recruitment period besides difficulties in recruiting
41
42 331 subjects with shoulder instability and laxity only thirteen subjects with an affected
43
44 332 shoulder were included. Naturally, this also affected the prevalence of positive and
45
46 333 negative test findings meaning that the prevalence of 0.50, as recommended in
47
48 334 reliability studies,²⁵ in all six tests was not accomplished. However, to overcome this,
49
50
51
52
53
54
55
56
57
58
59
60

Inter-tester reliability of clinical shoulder instability and laxity tests

1
2
3
4 335 PABAK calculations was used and reported along with kappa, to show transparently
5
6 336 how data would have been with equal distributions of positive and negative test
7
8
9 337 results. Nevertheless, future studies should use inclusion criteria of more established
10
11 338 shoulder instability and laxity conditions, and, if possible, verified by objective criteria
12
13 339 as surrogate for a gold standard of shoulder instability and laxity. This may optimize
14
15 340 prevalence as well as diagnostic accuracy in studies where this is a further aim.
16
17
18 341 The strengths of the study are the use of standardized procedures (including blinding
19
20 342 to patient status and the use of a three-phased protocol for conducting reliability
21
22 343 studies). Also, presentation of raw data, using contingency tables, along with kappa
23
24 344 and PABAK values, increases data transparency and improves interpretation of the
25
26 345 reliability study.
27
28
29
30

31 346 **Conclusions**

32
33 347 This study showed acceptable inter-tester reliability for four of six clinical shoulder
34
35 348 instability and laxity tests in relatively sports active individuals with and without self-
36
37 349 reported shoulder problems. However, relocation and sulcus sign tests need further
38
39 350 standardisation before being recommended for use in clinical practice. Based on the
40
41 351 frequency and mutual dependency of the current tests, especially apprehension and
42
43 352 surprise tests seem important in the characterisation of self-reported shoulder
44
45 353 problems. Future research on the validity of tests for shoulder instability and laxity is
46
47 354 needed.
48
49
50
51

52 355
53
54
55 356
56
57
58
59
60

Inter-tester reliability of clinical shoulder instability and laxity tests

1
2
3
4 357 **Author statement:** HE, KGI, CML and BJK conceived and designed the study. HE and
5
6 358 BHK recruited study participants and collected data. HE performed the statistical
7
8
9 359 analysis. HE, KGI, CML, BJK interpreted the results. HE drafted the manuscript with KGI,
10
11 360 CML, BHK and BJK contributing to the manuscript. All authors have read and approved
12
13 361 the final manuscript. HE is the guarantor.
14

15
16 362

17
18 363 **Consent for publication:** The depicted individuals in Table 1 have provided written
19
20 364 consent for publication of the pictures.
21

22
23 365

24
25 366 **Funding:** This work was supported by Region of Southern Denmark's Research fund
26
27 367 and The Danish Rheumatism Association.
28

29
30 368

31
32 369 **Competing interest:** None declared
33

34
35 370

36
37 371 **Data sharing statement:** No additional data are available
38

39
40 372
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

References

1. Luime, J. J., B. W. Koes, I. J. Hendriksen, A. Burdorf, A. P. Verhagen, H. S. Miedema, and J. A. Verhaar. 2004. 'Prevalence and incidence of shoulder pain in the general population; a systematic review', *Scandinavian Journal of Rheumatology*, 33: 73-81.
2. Farrar, N. G., J. J. Malal, J. Fischer, and M. Waseem. 2013. 'An overview of shoulder instability and its management', *Open Orthopaedics Journal*, 7: 338-46.
3. Goss, T. P. 1988. 'Anterior glenohumeral instability', *Orthopedics*, 11: 87-95.
4. Cameron, K. L., M. L. Duffey, T. M. DeBerardino, P. D. Stoneman, C. J. Jones, and B. D. Owens. 2010. 'Association of generalized joint hypermobility with a history of glenohumeral joint instability', *J Athl Train*, 45: 253-8.
5. Chahal, J., J. Leiter, M. D. McKee, and D. B. Whelan. 2010. 'Generalized ligamentous laxity as a predisposing factor for primary traumatic anterior shoulder dislocation', *J Shoulder Elbow Surg*, 19: 1238-42.
6. Dodson, C. C., and F. A. Cordasco. 2008. 'Anterior glenohumeral joint dislocations', *Orthop Clin North Am*, 39: 507-18, vii.
7. Jaggi, A., and S. Lambert. 2010. 'Rehabilitation for shoulder instability', *Br J Sports Med*, 44: 333-40.
8. Wilk, K. E., L. C. Macrina, and M. M. Reinold. 2006. 'Non-operative rehabilitation for traumatic and atraumatic glenohumeral instability', *N Am J Sports Phys Ther*, 1: 16-31.
9. Hegedus, E. J., A. P. Goode, C. E. Cook, L. Michener, C. A. Myer, D. M. Myer, and A. A. Wright. 2012. 'Which physical examination tests provide clinicians with the most value when examining the shoulder? Update of a systematic review with meta-analysis of individual tests', *Br J Sports Med*, 46: 964-78.
10. Ann Ganestam, Mikkel L. Attrup, Per Hølmich and Kristoffer W Barfoed. 2015. 'Evaluation of Clinical Practice of Shoulder Examination among Ten Experienced Shoulder Surgeons', *Journal of Orthopaedic Research and Physiotherapy*, 1.
11. Luime, J. J., A. P. Verhagen, H. S. Miedema, J. I. Kuiper, A. Burdorf, J. A. Verhaar, and B. W. Koes. 2004. 'Does this patient have an instability of the shoulder or a labrum lesion?', *JAMA*, 292: 1989-99.
12. Tzannes, A., and G. A. Murrell. 2002. 'Clinical examination of the unstable shoulder', *Sports Medicine*, 32: 447-57.
13. Gagey, O. J., and N. Gagey. 2001. 'The hyperabduction test', *J Bone Joint Surg Br*,

Inter-tester reliability of clinical shoulder instability and laxity tests

- 1
2
3
4 418 83: 69-74.
5 419
6 420 14. Lo, I. K., B. Nonweiler, M. Woolfrey, R. Litchfield, and A. Kirkley. 2004. 'An
7 421 evaluation of the apprehension, relocation, and surprise tests for anterior shoulder
8 422 instability', *Am J Sports Med*, 32: 301-7.
9 423
10 424 15. Silliman, J. F., and R. J. Hawkins. 1993. 'Classification and physical diagnosis of
11 425 instability of the shoulder', *Clin Orthop Relat Res*: 7-19.
12 426
13 427 16. Cavallo, R. J., and K. P. Speer. 1998. 'Shoulder instability and impingement in
14 428 throwing athletes', *Med Sci Sports Exerc*, 30: S18-25.
15 429
16 430 17. Matsen, F. A., 3rd, and J. D. Zuckerman. 1983. 'Anterior glenohumeral
17 431 instability', *Clin Sports Med*, 2: 319-38.
18 432
19 433 18. Robinson, C. M., J. Howes, H. Murdoch, E. Will, and C. Graham. 2006. 'Functional
20 434 outcome and risk of recurrent instability after primary traumatic anterior shoulder
21 435 dislocation in young patients', *J Bone Joint Surg Am*, 88: 2326-36.
22 436
23 437 19. van Kampen, D. A., T. van den Berg, H. J. van der Woude, R. M. Castelein, C. B.
24 438 Terwee, and W. J. Willems. 2013. 'Diagnostic value of patient characteristics,
25 439 history, and six clinical tests for traumatic anterior shoulder instability', *J Shoulder
26 440 Elbow Surg*, 22: 1310-9.
27 441
28 442 20. Warby, S. A., T. Pizzari, J. J. Ford, A. J. Hahne, and L. Watson. 2014. 'The effect of
29 443 exercise-based management for multidirectional instability of the glenohumeral
30 444 joint: a systematic review', *J Shoulder Elbow Surg*, 23: 128-42.
31 445
32 446 21. Gibson, K., A. Growse, L. Korda, E. Wray, and J. C. MacDermid. 2004. 'The
33 447 effectiveness of rehabilitation for nonoperative management of shoulder
34 448 instability: a systematic review', *J Hand Ther*, 17: 229-42.
35 449
36 450 22. Levy, A. S., S. Lintner, K. Kenter, and K. P. Speer. 1999. 'Intra- and interobserver
37 451 reproducibility of the shoulder laxity examination', *Am J Sports Med*, 27: 460-3.
38 452
39 453 23. Tzannes, A., A. Paxinos, M. Callanan, and G. A. Murrell. 2004. 'An assessment of
40 454 the interexaminer reliability of tests for shoulder instability', *J Shoulder Elbow
41 455 Surg*, 13: 18-23.
42 456
43 457 24. Kottner, J., L. Audige, S. Brorson, A. Donner, B. J. Gajewski, A. Hrobjartsson, C.
44 458 Roberts, M. Shoukri, and D. L. Streiner. 2011. 'Guidelines for Reporting Reliability
45 459 and Agreement Studies (GRRAS) were proposed', *J Clin Epidemiol*. 64 (1): 96-106.
46 460
47 461 25. Patijn J, Remvig L. . 2007. "Reproducibility And Validity 2007-Protocol Formats
48 462 For Diagnostics Procedures In Manual/Musculoskeletal Medicine." In.

Inter-tester reliability of clinical shoulder instability and laxity tests

- 1
2
3
4 463
5 464 26. Downie, W. W., P. A. Leatham, V. M. Rhind, V. Wright, J. A. Branco, and J. A.
6 465 Anderson. 1978. 'Studies with pain rating scales', *Annals of the Rheumatic Diseases*,
7 466 37: 378-81.
8 467
9
10 468 27. Kirkley, A., S. Griffin, H. McLintock, and L. Ng. 1998. 'The development and
11 469 evaluation of a disease-specific quality of life measurement tool for shoulder
12 470 instability. The Western Ontario Shoulder Instability Index (WOSI)', *Am J Sports*
13 471 *Med*, 26: 764-72.
14 472
15
16 473 28. 'World Medical Association Declaration of Helsinki: ethical principles for
17 474 medical research involving human subjects'. 2000. *JAMA*, 284: 3043-5.
18 475
19 476 29. Bahk, M., E. Keyurapan, A. Tasaki, E. L. Sauers, and E. G. McFarland. 2007.
20 477 'Laxity testing of the shoulder: a review', *Am J Sports Med*, 35: 131-44.
21 478
22
23 479 30. Sim, J., and C. C. Wright. 2005. 'The kappa statistic in reliability studies: use,
24 480 interpretation, and sample size requirements', *Phys Ther*, 85: 257-68.
25 481
26
27 482 31. Byrt, T., J. Bishop, and J. B. Carlin. 1993. 'Bias, prevalence and kappa', *J Clin*
28 483 *Epidemiol*, 46: 423-9.
29 484
30 485 32. Hoehler, F. K. 2000. 'Bias and prevalence effects on kappa viewed in terms of
31 486 sensitivity and specificity', *J Clin Epidemiol*, 53: 499-503.
32 487
33
34 488 33. Landis, J. R., and G. G. Koch. 1977. 'The measurement of observer agreement for
35 489 categorical data', *Biometrics*, 33: 159-74.
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Inter-tester reliability of clinical shoulder instability and laxity tests

490 **Figure legends**

491

492 Figure 1. Apprehension

493

494 Figure 2. Relocation

495

496 Figure 3. Surprise

497

498 Figure 4. Load and shift – anterior direction

499

500 Figure 5. Load and shift – posterior direction

501

502 Figure 6. Sulcus sign

503

504 Figure 7. Gagey

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



Table 1, Figure 1

26x19mm (300 x 300 DPI)

er review only



Table 1, Figure 2

26x19mm (300 x 300 DPI)

er review only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



Table 1, Figure 3

26x21mm (300 x 300 DPI)

review only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



Table 1, Figure 4

30x19mm (300 x 300 DPI)

er review only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



Table 1, Figure 5

28x23mm (300 x 300 DPI)

review only

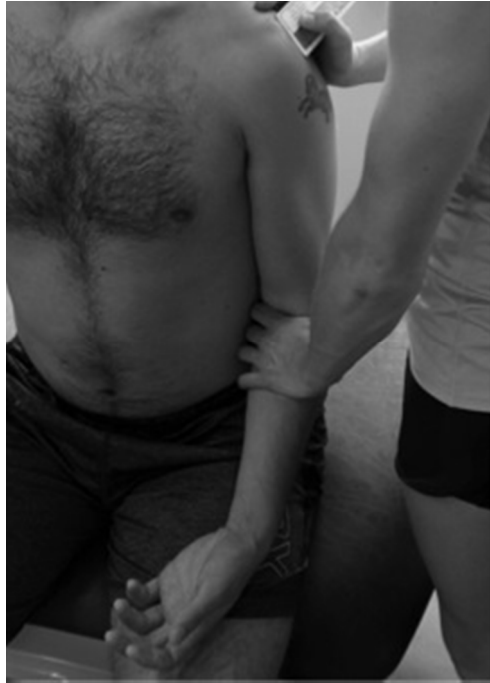


Table 1, Figure 6

20x28mm (300 x 300 DPI)

view only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

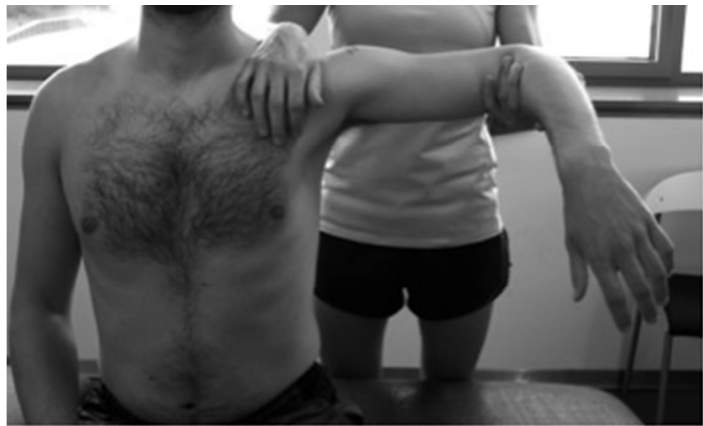


Table 1, Figure 7

29x17mm (300 x 300 DPI)

For peer review only