2024 SECEC Norbert Gschwend Prize

(Back) to the basics: diagnostic value of physical examination in elbow pathology

Elisa L. Zwerus, MD

Supervisors

Denise Eygendaal, MD, PhD

Michel P.J. van den Bekerom, MD, PhD

Bertram The, MD, PhD

Table of contents

Part I *Assessment of the range of motion of the elbow joint*

Chapter 1 **Normative values and affecting factors for the elbow range of motion**.

Chapter 2 **The reliability and validity of goniometric elbow measurements: a systematic review of the literature**

Chapter 3 **Validity and reliability of elbow range of motion measurements using digital photographs, movies and a goniometry smartphone application**

Part II *Physical examination of the elbow*

Chapter 4 **Physical examination of the elbow, what is the evidence? A systematic literature review**

Chapter 5 **Distal biceps tendon ruptures: diagnostic strategy through physical examination**

Chapter 6 **Clinical diagnosis of lateral sided elbow pain: predictors to recognize a diagnosis other than a tennis elbow**

Chapter 7 **Accuracy and quality of educational videos for elbow physical examination: A search from the earliest year until October 2018.**

Part I

Assessment of the range of motion of the elbow joint

Chapter 1

Normative values and affecting factors for the elbow range of motion

Elisa L. Zwerus, MD[1], Nienke W. Willigenburg, PhD[1], Vanessa A. Scholtes, PhD[1], Matthijs P. Somford, MD[2], Denise Eygendaal, MD, PhD[3], Michel P.J. van den Bekerom, MD[1]

[1] Shoulder and elbow unit, Department of orthopedic surgery, Onze Lieve Vrouwe Gasthuis, Amsterdam

[2] Department of orthopedic surgery, Rijnstate Hospital, Arnhem

[3] Upper limb unit, Department of orthopedic surgery, Amphia, Breda

Shoulder&Elbow 2017 sept 11. doi:10.1177/1758573217728711.

Abstract

Introduction. Abnormalities in the elbow range of motion (ROM) can be subtle, therefore it is important that the examiner can compare findings with reliable reference values, matching the patients' characteristics. Primary, we aimed to provide normative values for the elbow ROM in sub-populations based on age, gender, dominance, and body mass index (BMI). The secondary objective was to determine intra- and inter-rater reliability.

Methods. aROM (active range of motion) and pROM (passive range of motion) were measured bilaterally in healthy adults using a universal goniometer (UG). The influence of factors affecting the ROM was calculated using Pearsons Correlation Coefficient. In two samples of subjects, intra-rater and inter-rater reliability was determined.

Results. The study population (n=352) consisted of 47.2% male and 52.8% female subjects. For aROM (dominant hand), mean flexion was 146°, extension -2°, pronation 80° and supination 87°. Male subjects had smaller ROM compared to females (p<0.001). Differences

between dominant and non-dominant hand were less than one degree. pROMs were 3° to 5° larger than aROMs ($p<0.001$). Intra- and inter-rater reliability was good.
Conclusions. Elbow ROM is influenced by age, gender and BMI. In the general population, the ROM of the uninjured side can serve as a reference in case of an injured elbow.

Background

The elbow joint allows us to perform flexion-extension and pronation-supination movement. According to literature, values for flexion lie between 130° and 154° and extension between -6° and 11°. Pronation varied from 75° to 85° and supination from 80° to 104°[32, 50, 63, 74, 147, 158]. Daily activities can be performed with an elbow extension restriction of 30° and minimal flexion of 130 degrees, in combination with 50 degrees of pronation and supination[106], more extensive activities such as handling a cell phone require more mobility[124, 133]. Sports activities also require a greater elbow range of motion (ROM), however, athletes are also prone to reduced elbow ROM, especially overhead throwing athletes. Literature shows that the flexion-extension range is decreased by about 14° in asymptomatic baseball pitchers[158] and a significant decrease in passive elbow ROM in the first 24 hours after throwing[127]. Furthermore, overhead throwing athletes often also have a reduced shoulder ROM, which makes them prone for elbow injuries[138, 157].
Conversely, a restriction in elbow ROM will result in a greater load on the shoulder and wrist, causing injuries to those joints

Measuring the ROM is considered an integral part of physical examination for elbow pathology. Flexion is usually restricted by anterior soft tissues such as the biceps muscle, resulting in an elastic end-feeling. Pronation and supination also have an elastic end-feeling caused by ligaments, the interosseous membrane and forearm muscles[148]. Extension however has a hard end-feel due to the olecranon pushing into the humerus. The end-feeling provides information on the cause of the ROM restriction.
To accurately measure the elbow ROM, the Universal Goniometer (UG) is an easy, reliable and commonly used assessment method[5, 20, 32, 50, 56, 115, 129]. ROM assessment is important for both acute injuries, chronic injuries and to evaluate the effectiveness of a treatment. Athletes are more prone to acute elbow injuries, mostly fractures, for example in gymnasts

falling on an extended elbow. In traumatic injuries, elbow extension can be used as a sensitive clinical screening test for traumatic injuries. When the injured athlete has an unrestricted elbow extension, a fracture can be ruled out without an X-ray[42, 85]. Malunion, long period of immobilization and heterotopic ossifications following a traumatic injury are risk factors for elbow stiffness.

There are also a number of chronic overuse injuries, mostly in overhead athletes, resulting in ROM restrictions[72]. For example loose bodies, chondromalacia, valgus extension overload syndrome, osteochondritis dissecans in children and osteoarthritis in elderly can decrease elbow mobility[79]. Intra-articular fluid can also restrict motion, positioned in 70° of flexion, the pressure from intra-articular fluid and pain are lowest[104].

Apart from initial assessment of the elbow, measurement of the ROM is a key element to evaluate the effectiveness of a treatment for elbow stiffness.

Previous studies reported ROM measurements based on either small (under 50 subjects) or specific (e.g. athletes, injured or only one gender/dominant hand) study populations[32, 50, 63, 74, 147, 158]. Soucie et al. (2011) included a bigger study population, but did not account for factors such as BMI, hand dominance and only reported passive ROM[147].

In order to guide clinical decisions, it is important that reliable normative values are available and influencing factors are known. Therefore, the primary objective of our study was to provide normative values for subpopulations based on age, gender, hand dominance and BMI. Also, correlations between ROM and age and BMI will be determined. Next to normative values for subpopulations, we aimed to provide standard values for *passive* ROM (pROM) and *active* ROM (aROM) of the elbow (i.e. flexion, extension, pro- and supination) for the total population.

Our secondary objective was to define intra-rater and inter-rater reliability of goniometric measurement of the elbow ROM and present this in a comprehensible way, to facilitate interpretation of changes observed in the clinic.

Methods

Study design

The present study assessed the ROM using a cross-sectional design. Data was collected from August 2015 to October 2015 and researchers were based in OLVG hospital, Amsterdam, the Netherlands. Ethical approval was waived by the local ethical committee (WO15.069).

Study population

Subjects in the age of 18 to 79 years old who volunteered to undergo a ROM examination were included. Subjects diagnosed with a disease or previous injury that could potentially affect elbow ROM on either right or left side were excluded. Volunteers were recruited at (pre-informed) public spaces and events, such as sports clubs/events, businesses, schools/universities, family/friends gatherings, supermarkets etc. Effort was made to include a minimum of 25 subjects in each age category as following: 18-29, 30-39, 40-49, 50-59, 60-69 and 70-79 years old.

Data collection

All subjects were examined by a clinician in the orthopaedic department (EZ), trained for elbow ROM measurements using a predefined protocol developed by two human movement scientists (NW, VS), an orthopaedic surgeon (MB) and a physiotherapist. For flexion and extension measurement the acromion and radial styloid process were used as landmarks for the goniometers' arms and the lateral epicondyle as the centre of rotation. Flexion was measured with 0° shoulder anteflexion and 0° abduction and a maximally supinated forearm. With the same forearm position and the shoulder in 90° anteflexion and 0° abduction the extension was measured. In both measurements, the centre of rotation for pronation and supination was the ulnar styloid process and the goniometers' arms were placed parallel to the humeral midline and dorsal or volar wrist respectively. Both measurements were taken with 90° of elbow flexion and manual fixation of the upper arm to the body. The shoulder and elbow position and landmarks for the goniometer positioning were based on guidelines and recommendations from previous literature[27, 36, 50, 52, 117, 137, 140, 150]. aROM was measured by asking the patient to make the movement as far as they could, pROM was measured with added support by the examiner in the direction of the movement.

aROM and pROM were measured in both arms of each subject. Execution of ROM measurements with landmarks are presented in figures 1A-D.

Figure 1A – Flexion

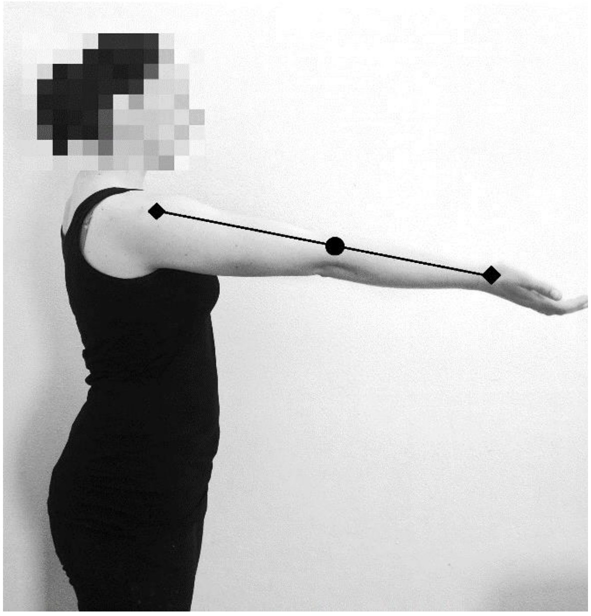


Figure 1B – Extension

Figure 1C – Pronation

Figure 1D – Supination

Of each subject age (years), gender (male/female), height (centimetres), weight (kilograms) and hand dominance (right/left) were obtained. Subjects were not asked for any personal details and data was stored using anonymous consecutive study numbers.

For intra-rater reliability analysis, 20 subjects' both elbows' range of motion was measured twice by the primary examiner within a time frame between 1 to 7 days. To analyse the inter-rater reliability, 10 subjects' both elbows were measured by the primary examiner and a trained orthopaedic surgeon, blinded for the first measurement results. The first measurement by the primary examiner of these 30 subjects were also included for the final study. The subsequent 322 subjects were assessed once.

Statistical analysis

All data were collected using Microsoft Excel 2010 and analysed using Statistical Package for Social Science for Windows (SPSS Statistics version 22, Armonk, NY, USA: IBM Corp). Before analysing, all data was double-checked by visual inspection and box-plots. A p-value of <0.05 was considered statistically significant.

Normative values based on subgroups. An unpaired samples t-test was used to compare patients' demographics and to compare ROM measurements between gender, age categories and BMI categories. Differences for dominant and non-dominant arm were tested with a paired samples t-test. Subjects were divided in different age groups: 18-29, 30-39, 40-49, 50-59, 60-69 and 70-79 years old. BMI was divided in categories <18.5 (underweight), 18.5-24.9 (normal), 25.0-29.9 (overweight) and ≥30.0 (obese)[1].

Pearson's correlation coefficients were calculated to determine the influence of the continuous parameters BMI and age on the elbow range of motion.

Intra-rater and inter-rater reliability analysis. The ICC and 95% confidence intervals (CI) were calculated for each ROM measure using a two-way random effects model where both people effects and measures effects are random. ICC of >0.75 indicate good reliability[121]. In addition to the ICC's, for the contribution of variance caused by subjects ($Var_{subject}$), occasion ($Var_{occasion}$) or measurement error (Var_{error}) was determined using variance components analysis, in order to calculate the Standard Error of Measurement (SEM) and the Smallest Detectable Difference (SDD) in Excel. SEM was calculated using the following formula: $SEM = \sqrt{(Var_{occasion} + Var_{error})}$ and SDD using the following formula: $SDD = \sqrt{2} * 1.96 * \sqrt{(Var_{occasion} + Var_{error})}$[128].

Results

Subject demographics

The distribution of age, hand dominance and BMI by gender of the 352 participants is presented in table 1. Unpaired t-test did not show a statistically significant difference between the demographics of the male and female subjects (p<0.05).

		Male (n=166, 47.2%) N (%)	Female (n=186, 52.8%) N (%)
Age in years, mean (95% CI)		49 (45 to 51)	45 (42 to 47)
Age categories (years)			
	18-29	33 (19.9%)	57 (30.6%)
	30-39	27 (16.3%)	25 (13.4%)
	40-49	28 (16.9%)	29 (15.6%)
	50-59	27 (16.3%)	25 (13.4%)
	60-69	25 (15.1%)	25 (13.4%)
	70-79	26 (15.7%)	25 (13.4%)
BMI mean (95% CI)		24.2 (23.6 to 24.7)	23.5 (23.0 to 24.1)
BMI categories			
	<18.50	6 (3.6%)	8 (4.3%)
	18.50-24.99	105 (63.3%)	133 (71.5%)
	25.00-29.99	40 (24.1%)	34 (18.3%)
	≥30.00	15 (9.0%)	11 (5.9%)
Dominance			
	Right	150 (90.4%)	167 (89.8%)
	Left	16 (9.6%)	19 (10.2%)

Table 1 - Subject demographics compared by gender (n=352)

Normative values for ROM based on subgroups

Table 2 presents average aROM and pROM values for male and female subjects (dominant and non-dominant side separately). Unpaired t-test showed significantly smaller ROM for all measurements for male subjects compared to females (p<0.01).

Movement		ROM in degrees, mean (95% CI)			
		Male (n=166)		Female (n=186)	
		Dominant	Non-dominant	Dominant	Non-dominant
Active	Flexion	143 (142 to 144)	143 (142 to 144)	148 (147 to 149)	148 (147 to 149)
	Extension	0 (0 to 1)	0 (-1 to 1)	-5 (-6 to -4)	-5 (-6 to -5)
	Pronation	78 (77 to 79)	79 (78 to 80)	82 (81 to 83)	83 (82 to 84)
	Supination	85 (84 to 86)	85 (83 to 86)	88 (87 to 89)	88 (87 to 89)
Passive	Flexion	147 (146 to 148)	147 (146 to 148)	153 (152 to 154)	153 (152 to 154)
	Extension	-2 (-2 to -1)	-2 (-3 to -1)	-8 (-9 to -7)	-9 (-10 to -8)
	Pronation	82 (81 to 83)	83 (81 to 84)	86 (85 to 87)	87 (86 to 88)
	Supination	89 (88 to 90)	88 (87 to 90)	93 (92 to 94)	92 (9 to 93)

Table 2 – Normative values for elbow ROM by gender

Differences in ROM between dominant and non-dominant side varied from 0.3 to 0.7 degrees and were statistically significant (p<0.05) for all movements, except for active flexion. For all movements except pronation the ROM of the dominant side was slightly larger compared to the non-dominant side.

Compared to aROM, pROM was higher in flexion, pronation and supination and more negative in extension, all statistically significant (p<0.001). Tables 3 and 4 present normative values by age and BMI, with aROM and pROM separately, for the dominant side. Differences between age groups (by gender) were all statistically significant (p<0.001). For BMI, in females all measurements between groups were statistically significant (p<0.001). In males active and passive flexion and passive pronation showed a statistically significant difference between BMI groups (p<0.05). Note that the underweight and obese groups are relatively small.

Dominant side ROM in degrees, mean (95% CI)						
Age category (years)	18-29	30-39	40-49	50-59	60-69	70-79
Male (n=166)	33	27	28	27	25	26

Active	Flexion	146 (144 to 148)	145 (142 to 147)	143 (141 to 146)	143 (141 to 146)	141 (139 t 143)	139 (138 to 141)
	Extension	-1 (-3 to 0)	-1 (-2 to 0)	0 (-1 to 2)	0 (-1 to 1)	1 (-1 to 3)	2 (1 to 4)
	Pronation	81 (79 to 83)	80 (77 to 83)	81 (78 to 83)	79 (77 to 81)	74 (72 to 76)	74 (72 to 76)
	Supination	89 (88 to 91)	87 (84 to 90)	86 (84 to 88)	85 (83 to 87)	81 (78 to 84)	81 (79 to 83)
Passive	Flexion	151 (149 to 153)	150 (147 to 152)	147 (145 to 149)	148 (146 to 150)	144 (142 to 146)	142.7 (141 to 144)
	Extension	-3 (-5 to -2)	-2 (-4 to -1)	-2 (-4 to 1)	-2 (-3 to 0	0 (-2 to 1)	1 (-1 to 2)
	Pronation	86 (84 to 87)	83 (80 to 86)	85 (82 to 87)	83 (81 to 85)	78 (75 to 80)	77 (75 to 79)
	Supination	94 (92 to 96)	92 (89 to 95)	91 (89 to 93)	89 (87 to 91)	84 (82 to 87)	83 (81 to 85)
Female (n=186)		57	25	29	25	25	25
Active	Flexion	149 (148 to 151)	153 (152 to 155)	149 (147 to 151)	148 (146 to 149)	145 (143 to 148)	142 (140 to 144)
	Extension	-8 (-10 to -6)	-6 (-8 to -4)	-5 (-7 to -3)	-1 (-3 to 2)	-2 (-4 to -1)	-2 (-4 to -01)
	Pronation	83 (82 to 85)	87 (85 to 89)	86 (84 to 87)	80 (77 to 82)	78 (75 to 81)	76 (74 to 78)
	Supination	92 (90 to 93)	93 (91 to 95)	90 (89 to 91)	87 (85 to 89)	84 (81 to 86)	81 (79 to 84)
Passive	Flexion	155 (154 to 157)	158 (156 to 160)	154 (152 to 156)	152 (150 to 154)	149 (146 to 152)	145 (143 to 147)
	Extension	-12 (-14 to -10)	-10 (-12 to -8)	-8 (-11 to -6)	-3 (-6 to 0	-5 (-7 to -4)	-4 (-6 to -3)
	Pronation	89 (88 to 91)	91 (90 to 93)	90 (89 to 91)	84 (80 to 86)	82 (79 to 85)	79 (76 to 81)

	Supination	96 (96 to 98)	98 (96 to 101)	94 (92 to 95)	91 (89 to 93)	87 (85 to 89)	84 (82 to 86)

Table 3 – Normative values by age

Dominant side ROM in degrees, mean (95% CI)					
BMI category		<18.50	18.50-24.99	25.00-29.99	≥30.00
Male (n=166)		6	105	40	15
Active	Flexion	151 (148 to 155)	145 (144 to 146)	140 (139 to 142)	136 (134 to 138)
	Extension	-2 (-4 to 1)	0 (-1 to 1)	1 (0 to 2)	0 (-3 to 2)
	Pronation	84 (80 to 89)	78 (77 to 80)	77 (75 to 79)	78 (75 to 81)
	Supination	90 (85 to 94)	86 (85 to 87)	83 (81 to 85)	82 (80 to 85)
Passive	Flexion	155 (152 to 159)	149 (148 to 150)	144 (143 to 1456)	140 (138 to 143)
	Extension	-3 (-8 to 2)	-2 (-3 to -1)	-1 (-2 to 0)	-2 (-4 to 1)
	Pronation	89 (85 to 92)	82 (81 to 84)	81 (78 to 83)	82 (79 to 85)
	Supination	95 (89 to 100)	90 (89 to 91)	87 (85 to 89)	86 (83 to 89)
Female (n=186)		8	133	34	11
Active	Flexion	157 (154 to 159)	149 (148-150)	146 (144 to 148)	138 (135 to 141)
	Extension	-15 (-17 to -12)	-5 (-6 to -4)	-3 (-6 to -1)	-2 (-5 to 2)
	Pronation	86 (82 to 89)	82 (81 to 84)	80 (78 to 82)	78 (75 to 82)
	Supination	94 (90 to 97)	89 (88 to 90)	8 (84 to 89)	84 (81 to 88)
Passive	Flexion	163 (161 to 166)	154 (153 to 155)	150 (148 to 152)	142 (139 to 145)
	Extension	-20 (-24 to -17)	-8 (-9 to -7)	-6 (-8 to -2)	-4 (-8 to -1)
	Pronation	91 (88 to 95)	87 (86 to 88)	84 (81 to 87)	81 (77 to 85)
	Supination	100 (95 to 105)	93 (92 to 95)	90 (88 to 92)	88 (84 to 92)

Table 4 – Normative values by BMI

Correlations

Age correlated moderately (r ≥ 0.50) negative with passive pronation (dominant -0.50; non-dominant -0.56) and supination (-0.56; -0.50). BMI correlated moderately negative on flexion, both active (-0.57; -0.59) and passive (-0.59; -0.59). Age correlated moderately negative with passive flexion (-0.59; -0.59). All other combinations showed low correlations (r < 0.50), ranging from -0.29 to -0.49 for flexion/pronation/supination and 0.22 to 0.44 for extension. No major differences in correlations were observed between dominant and non-dominant side.

Intra- and inter-rater reliability

In the subgroup (n=20) for intra-rater reliability analysis only the difference in age between the intra-rater subgroup and total group was statistically significant. Intra-rater ICCs for all measurements showed good reliability. SEM and SDD were higher for pronation and supination compared to flexion and extension.

Demographics of the subgroup (n=10) for inter-rater reliability were not statistically significant different compared to the total group. Inter-rater ICCs showed good reliability (>0.75) apart from passive flexion (0.74). SEM and SDD were slightly higher for pronation and supination compared to flexion and extension. Inter-rater ICC's, SEM and SDD were similar to intra-rater reliability. For both inter-rater and intra-rater reliability results from the dominant side were similar to non-dominant side. Intra-rater and inter-rater reliability results from the dominant side are presented in table 5.

Movement (dominant side)		Intra-rater reliability			Inter-rater reliability		
		ICC (95% CI)	SEM	SDD	ICC (95% CI)	SEM	SDD
Active	Flexion	0.76 (0.47-0.90)	3˚	7˚	0.86 (0.53-0.96)	2˚	5˚
	Extension	0.92 (0.80-0.97)	2˚	6˚	0.89 (0.63-0.97)	1˚	4˚
	Pronation	0.90 (0.77-0.96)	3˚	8˚	0.92 (0.47-0.98)	3˚	7˚
	Supination	0.91 (0.78-0.96)	3˚	8˚	0.87 (0.56-0.97)	3˚	8˚
Passive	Flexion	0.74 (0.45-0.89)	3˚	7˚	0.79 (0.33-0.94)	2˚	6˚

	Extension	0.95 (0.88-0.98)	2˚	5˚	0.85 (0.52-0.96)	2˚	5˚
	Pronation	0.86 (0.69-0.94)	3˚	9˚	0.91 (0.54-0.98)	3˚	8˚
	Supination	0.90 (0.75-0.96)	3˚	7˚	0.82 (0.46-0.95)	3˚	9˚

Table 5 – Intra- and interrater reliability - As an example, in dominant side active flexion the SEM was 3°, which means individual scores had an average measurement error of 3°. SDD in this measurement was 7°, which means that if a subjects' dominant side active elbow flexion changes more than 7°, the change is considered as a true change with 95% confidence.

Discussion

In our study normative values for the elbow range of motion derived from a population of 352 healthy volunteers are presented. For male subjects, all ROM measurements were statistically significant lower compared to females. Higher age correlated moderately with lower ROM for active supination and passive pronation and supination. A higher BMI correlated with a lower ROM for active and passive flexion. Differences between dominant and non-dominant side (<1°) were too small to be clinically relevant. Therefore, in a general population, the examiner is able to decide whether the measured ROM is abnormal or not, by comparing the injured side with the uninjured side. This of course combined with the information on the intra-rater reliability and patient specific characteristics such as (overhead throwing) sport activities. If abnormality arises bilaterally, the normative values for specific subgroups (provided in tables 3 and 4) and correlations (in text) can be used to make an estimation of the expected ROM. Exact estimation of the ROM with all affecting factors incorporated such as gender, age and BMI is not possible due to the small numbers of subjects in certain groups (f.e. underweight elderly).

This study is not without limitations. A possible source of error might be unjustified inclusion of subjects who did not report factors that may influence the elbow ROM despite our written and oral inquiries. Given the large study population, we expect the influence of this potential source of error to be small. Furthermore, we did not ask for (professional) sports activities, therefore it is not possible to draw conclusions on the influence of sports on the elbow ROM.

We have chosen to exclude subjects under 18 years and therefore we could not draw conclusions for the paediatric population. Golden et al. (2007) published a study presenting mean elbow ROM in 300 healthy children supplemented with age- and gender related changes[58].

In previous literature, two studies investigated differences between the ROM of the dominant and non-dominant side. One study showed statistically significant differences, with mean difference varying from 1.7° to 2.6°[63] and another study showed non-significant differences[95]. Similar to our study, in both studies differences were too small to be considered clinically relevant. However, in overhead throwing athletes, the flexion-extension range may be greatly (about 14°) decreased[158], even more distinct up to 24 hours after throwing[127]. A study on the effect of obesity on active flexion showed no significant differences compared to subjects with an healthy BMI (20-25 kg/m^2)[114], in contrast to our study. Nevertheless, significant negative correlations for BMI on elbow range of motion (flexion/extension) were found in obese children (right r = -0.54, left r = -0.43)[59].

For measuring elbow ROM, the universal goniometer is considered to be the most easy to use and clinically available tool[32, 57, 68, 74, 153]. Studies published in the past show a good intra-rater and inter-rater reliability of goniometric elbow measurements for all measurements in both injured and healthy volunteers, with ICC ranging from 0.70 to 0.99 [20, 32, 50, 56, 129]. However, a study by Armstrong et al. [5] showed lower inter-rater and intra-rater reliability in injured subjects for flexion and extension (0.45-0.99). In handball players, lower intra-rater reliability for flexion and extension were observed (0.49-0.93)[50]. One small study about inter-rater reliability in healthy subjects showed disappointing results for flexion and extension (ICC 0.53)[115]. Because the ICC uses variance between subjects' ROM measurements to calculate reliability, a large variation between subjects will lead to a higher ICC. This could possibly draw a misleading conclusion of good reliability. Therefore we decided in our study to provide the statistical measures SEM and SDD. These measures provide more meaningful information for repeated measures within subjects, which is relevant for clinical practice [128, 135]. One study reported results in degrees for intra-rater reliability (mean within 1.2°) and inter-rater reliability (mean within 1.6°)[63]. Another study presented intra-rater SDD's of 7.8° for active flexion and 6.3° for active extension and inter-rater SDD's of 8.2° for flexion and 6.3° for extension[36]. Intra-rater SEM was reported in one study with results from 1.4° to 1.6° for flexion and 1.0° to 1.2° for extension[50]. No previous studies were found that reported SEM or SDD values for pronation and supination.

In our reliability analysis, both intra-rater and inter-rater ICCs for all measurements showed good reliability, ranging from 0.74-0.95. Reliability analysis for dominant and non-dominant side was not statistically significant different, therefore we presented dominant side only. Relative low ICC's were found for flexion, due to the small variation of flexion ROM amongst subjects. Our results corresponded with studies measuring non-injured subjects[20, 32, 50, 56, 115]. Intra-rater SEM values ranged from 2° to 3° and SDD values ranged from 5° to 9°. Of these measures, passive pronation was the least reliable with a SEM of 3° and SDD of 9°. So in this example, an individual score had an average measurement error of 3° and for a repetitive measurement, the difference must be minimum 9° to be considered as a true change with 95% confidence. Inter-rater SEM ranged from 1° to 3° and SDD from 4° to 9° and applies to measurements between different examiners.

Conclusion

Elbow ROM measurement was conducted on 352 healthy subjects and was found to be influenced by gender, age and BMI. For all movements, males have a significant smaller range of motion. Also, a higher age correlates with a smaller passive pronation-supination range of motion and a higher BMI with less flexion. Differences between dominant and non-dominant side were too small to be clinically relevant, therefore in the general population the ROM of the opposite side could serve as a reference for the injured side.

Chapter 2

The reliability and validity of goniometric elbow measurements: a systematic review of the literature

Elisa L. Zwerus, MD[2], Suzanne F. van Rijn, MD[1], Koen L.M. Koenraadt, PhD[1]

Wilco C.H. Jacobs, PhD[1], Michel P.J. van den Bekerom, MD[2], Denise Eygendaal, MD, PhD[1,3]

1. Orthopedic surgery department, Amphia hospital, Breda, Netherlands
2. Department of orthopedic surgery, Shoulder and elbow unit, Onze Lieve Vrouwe Gasthuis, Amsterdam
3. Department of orthopedic surgery, Academic Medical Centre, Amsterdam

Shoulder & Elbow 2018 june 3. Doi: 10.1177/2F1758573218774326

Abstract

Background: The universal goniometer (UG) is a simple measuring tool. With this review we aimed to investigate the reliability and validity of the UG in measurements of the elbow.

Methods: PRISMA guidelines were followed and our study protocol was published online at PROSPERO. A literature search was conducted on relevant studies using Embase, Medline ovid, Web of science, Scopus, Cochrane, Pubmed publisher, Cinahl EBSCO and Google scholar. Methodological quality was assessed using the QAREL scoring system.

Results: Out of 697 studies yielded from our literature search, 12 were included. Six studies were rated as high quality. The intrarater reliability intraclass correlation coefficient (ICC) ranged from 0.45-0.99, the interrater reliability ranged from ICC 0.53-0.97. One study providing instructions on goniometric alignment did not find a difference in expert versus non-expert examiners. Another study in which examiners were not instructed, found a higher interrater reliability in expert examiners. One study investigating the validity of the

goniometer in elbow measurements, found a maximum standard error of the mean of 11.5° for total range of motion.

Discussion: Overall, the studies showed high intra- and interrater reliability of the UG. The reliability of the UG in non-expert examiners can be increased by clear instructions on goniometric alignment.

Background

A patients' ability to perform daily activities such as eating, combing hair, writing and using a PC is highly dependent on the range of motion (ROM) of the elbow. A restricted elbow ROM can result in a serious disability. According to literature on the elbow ROM, in a healthy person values for flexion lie between 130° and 154° and extension between -6° and 11°. Pronation varied from 75° to 85° and supination from 80° to 104° [32, 50, 63, 74, 147, 158, 160]. A decrease in ROM can be an indicator of chronic and progressive pathology, such as heterotopic ossifications, osteoarthritis, loose bodies, chondromalacia, valgus extension overload syndrome and osteochondritis dissecans [72, 79]. But also in traumatic injuries assessment of elbow ROM can be important, for example an unrestricted elbow extension can rule out a fracture without an X-ray [42, 85].The degree of limitation in elbow ROM can also be used as an indicator of the impact of the disease in daily activities. Daily activities can be performed with an elbow extension restriction of 30° and minimal flexion of 130 degrees, in combination with 50 degrees of pronation and supination [106], however some activities as handling a cell phone require more mobility [124, 133]. Probably the most important reason to measure the elbow ROM is to closely monitor disease progression or treatment effectiveness.

The universal goniometer (UG) is a simple measuring tool, which is frequently used by many different health care professionals such as physiotherapists, general practitioners and orthopaedic surgeons. Other less studied modalities to measure the elbow ROM include the use of photography, movies, a smartphone application and visual estimation [20, 48, 100, 108, 113, 155]. To appreciate the "true value" of measurements, using the universal goniometer, it is important to know its' validity and intra- and interrater reliability.

General assumption is that the reliability of the UG is higher when used by an experienced tester and interrater variations are smallest when a standardized measuring method is used

[5, 52]. Many physicians and physiotherapists use the UG without using an identical measuring protocol if any available.
This systematic literature review investigated the reliability and validity of the UG in the measurement of elbow range of motion.

Methods

This systematic review was conducted according to the PRISMA statement (Preferred Reporting Items for Systematic reviews and Meta-Analysis) [103]. The study-protocol was published online at the PROSPERO International prospective register of systematic reviews (http://www.crd.york.ac.uk/PROSPERO) under registration number CRD42016043760.

Search strategy

A comprehensive electronic literature search was conducted in collaboration with an experienced clinical librarian on relevant articles published from the earliest year until October 2017 in the following databases: EMBASE, MEDLINE ovid, Web of science, Scopus, Cochrane, PubMed publisher, Cinahl EBSCO and Google scholar. The search terms (and synonyms) were 'elbow', 'goniomet', 'range of motion', 'accuracy', 'reliability', 'validity', and 'inter/intra observer'. (Appendix 1). The reference lists of included articles were manually checked for potentially relevant articles.
Only studies investigating the reliability and/or validity of the UG in elbow measurements in human adults were included. Range of motion included flexion, extension, pronation, supination and/or carrying angle. Exclusion criteria were: a language other than English or Dutch, subjects under the age of 18, animals, full text not available, a different measuring tool than the universal goniometer.

Study selection

Two authors (SR and WJ) independently screened all titles ad abstracts yielded by the search to identify relevant studies meeting the inclusion and exclusion criteria. The authors were not blinded for author and affiliation names of these studies. Then both authors assessed the full text of the selected articles. Afterwards, the reviewers compared their results, in case of differences they discussed until agreement was reached.

Quality assessment

The quality of the included articles was assessed by two authors (SR and KK) using the QAREL scoring system [91]. This tool scores the articles in their sampling bias: (1) the representativeness of subjects and raters, (2) rater blinding, (3) order of examination, (4) suitability of the time interval, (5) applied and interpreted appropriately and (6) statistical analysis. The maximum score is eleven. A study was considered having a high quality when it scored >60% and low quality when scored <60%. This cut-off point has been used in several previous studies [15, 61, 97].

Data extraction

The following data was extracted from all of the included articles by three authors independently (SR, EZ and KK): (1) population (healthy subjects or symptomatic patients); (2) number of subjects included; (3) movement measured (flexion, extension, pronation, supination or carrying angle); (4) active or passive range of motion; (5) if the measurement protocol was described; (6) if bony landmarks were used or defined prior to the measuring; (7) validity; (8) intra- and inter-observer reliability; and (9) information about the examiner (profession and/or level of experience in goniometry).

Data analysis

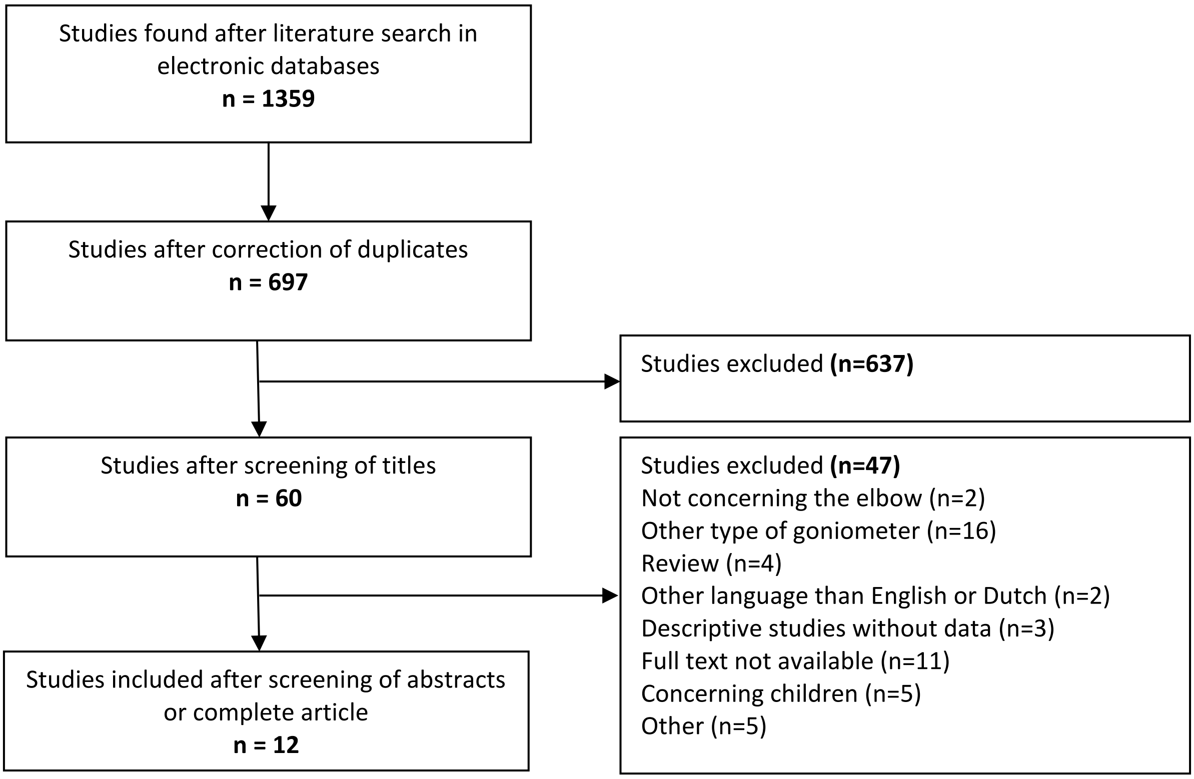
Data-analysis was performed by two independent authors (SR and EZ) using Microsoft Excel 2010 (Microsoft Corp. Washington, USA). ICC less than 0.40 was considered poor, between 0.40 and 0.59 fair, between 0.60 and 0.74 good and >0.75 excellent [34].

Results

Study selection

A total of 1386 articles were found. After removal of duplicates 697 articles remained. The titles of the 697 articles were screened and 60 articles were selected as potentially relevant. After reviewing abstracts and/or full text, 48 articles were excluded for various reasons, such as: review articles, subjects were children, full text not available, a language other than English or Dutch or the use of a measuring device other than the UG. Twelve articles were finally included for data extraction [5, 19, 32, 35, 50, 55, 56, 60, 90, 115, 129, 160]. Figure 1 shows the PRISMA flow chart.

Figure 1 PRISMA flow chart of study selection



Quality assessment

The QAREL checklist showed a high quality (score >60%) in six out of twelve studies [5, 19, 35, 55, 129, 160], all other studies were of low quality. Most of the studies rated as low quality did not blind (or did not mention to blind) the raters to the findings of other raters [32, 50, 56, 60, 90] or their own prior findings [5, 32, 35, 50, 60]. An overview of all the QAREL scores is presented in table I.

Table I - QAREL scores

		Armstrong 1998	Blonna 2012	Chapleau 2011	Cimatti 2013	Fieseler 2015	Flowers 2001	Gajdosik 2001	Goodwin 1992	Low 1976	Petherick 1988	Rothstein 1983	Zwerus 2017
1	Was the test evaluated in a sample of subjects who were representative of those tho whom the authors intended the results to be applied?	1	1	1	1	1	1	1	1	1	1	1	1
2	Was the test performed by raters who were representative of those to whom the authors intended the results to be applied?	1	1	1	0	1	1	Uncl	1	1	Uncl	1	1
3	Were raters blinded to the findings of other raters during the study?	1	1	0	1	N/A	1	N/A	Uncl	Uncl	1	1	1
4	Were raters blinded to their own prior findings of the test under evaluation?	0	1	0	0	Uncl	1	1	Uncl	1	1	1	1
5	Were raters blinded to the results of the accepted reference standard or disease status for the targeted disorder (or variable) being evaluated?	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
6	Were raters blinded to clinical information that was not intended to be provided as part of the testing procedure or study design?	1	Uncl	N/A	Uncl	Uncl	Uncl	N/A	N/A	N/A	N/A	Uncl	N/A
7	Were raters blinded to additional cues that were not part of the test?	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
8	Was the order of examination varied?	1	0	0	1	0	1	Uncl	Uncl	0	Uncl	Uncl	1
9	Was the stability of the variable being measured taken into account when determining the suitability of the time-interval between repeated measures?	1	1	N/A	1	1	1	1	1	N/A	N/A	1	1
10	Was the test applied correctly and interpreted appropiately?	1	1	1	1	1	1	1	0	1	1	1	1
11	Were appropriate statistical measures of agreement used?	1	1	1	1	1	Uncl	1	0	0	1	1	1
	Total	**8**	**7**	**4**	**6**	**5**	**7**	**5**	**3**	**4**	**5**	**7**	**8**
	score (%)	<u>73</u>	<u>64</u>	36	<u>55</u>	45	<u>64</u>	45	27	36	45	<u>64</u>	<u>73</u>

Included studies

Three studies tested the universal goniometer on symptomatic patients [5, 55, 129], seven studies used healthy volunteers [32, 50, 56, 60, 90, 115, 160], two studies included both healthy subjects and symptomatic patients [19, 35]. Together a number of 376 participants were included. A study by Low et al. [90] only included one subject, however this study used 50 raters. Rothstein et al. [129] used 12 raters with 12 subjects, all other studies used 5 or less raters with 23 up to 50 subjects. The number of measurements in all studies was two or three, the interval varied from consequently to four weeks apart. The age varied from 18 to 85 years. Nine studies tested elbow flexion, eight studies extension, 5 studies pronation and supination. Chapleau et al. also added the carrying angle [32]. Nine studies performed the measurements during active ROM, two during passive ROM [55, 129] and one study measured both active and passive [160]. In one study the arms of the subjects were in a fixed position [90]. In four studies the bony landmarks used for the measurements were defined [32, 60, 115, 160]. Two studies investigated the difference between expert and non-expert examiners. Armstrong et al. found no differences between expert and non-expert examiners[5], Blonna et al. found a slightly lower reliability in non-expert examiners [19]. Cimatti et al. compared injured to non-injured subjects, showing similar inter-rater reliability for pronation and supination[35]. Characteristics of included studies are presented in table II.

	Population			Measurements						Intra-rater		Inter-rater	
Study	N	Healthy/ symptomatic	Age (years)	Active/ Passive	Fl	Ex	Pro	Sup	CA	Trials (n)	Interval	Raters (n)	Expert/ non-expert
Armstrong 1998	38	Symptomatic	N/A	Active	+	+	+	+	-	2	Same day	5	Both
Blonna 2012	50	Both	18-85	Active	+	+	-	-	-	N/A	N/A	4	Both
Chapleau 2011	51	Healthy	19-50	Active	+	+	-	-	+	3	Consequently	2	N/A
Cimatti 2013	33	Both	18-70	Active	-	-	+	+	-	3	Consequently	2	Expert
Fieseler 2015	47	Healthy	18-25	Active	+	+	-	-	-	3	Week	N/A	N/A
Flowers 2001	30	Symptomatic	21-79	Passive	-	-	+	+	-	2	Same day	3	Expert
Gajdosik 2001	31	Healthy	19-40	Active	-	-	+	+	-	3	Consequently	1	N/A
Goodwin 1992	23	Healthy	18-31	Active	+	+	-	-	-	2	Month	3	Expert
Low 1976	1	Healthy	N/A	Active	+	-	-	-	-	N/A	N/A	50	Expert
Petherick 1987	30	Healthy	20-28	Active	+	+	-	-	-	3	Consequently	2	N/A
Rothstein 1983	12	Symptomatic	N/A	Passive	+	+	-	-	-	2	Same day	12	Expert
Zwerus 2017	30	Healthy	18-79	Both	+	+	+	+	-	2	Week	2	Expert
Total	376				9	8	5	5	1	25		86	

Table II – Study characteristics

Statistical analysis of results

Our intention was to perform heterogeneity analysis and, if applicable, meta-analysis on the included studies. Most studies use the intra-class coefficient (ICC) to express the inter-rater and intra-rater reliability. There are several different methods to compute the ICC, for example measuring ICC on single or average values [17, 18]. We attempted to determine the method of ICC calculation in every study, however this was not clear in all studies. Also, some studies presented a ICC range instead of a fixed number. Pooling results is inappropriate in this case[70]. Besides statistical heterogeneity, clinically the studies were also very heterogeneous. Therefore we decided to review the ICC's narratively.

Validity

One study investigated the validity of goniometric elbow measurements [32]. They compared the goniometric measurements of the elbow by one examiner with radiographic measurements of 51 healthy volunteers by two examiners. They found maximal errors of the goniometric measurements of 10.3° for extension, 7.0° for flexion, 11.5° for total ROM and 6.5° for the carrying angle.

Intra-rater reliability

Six studies investigated the intra-rater reliability for flexion [5, 32, 50, 60, 129, 160], showing fair to excellent reliability. Results for expert and non-expert raters [5] and passive and active measurements [160] were similar. Zwerus et al. calculated a Standard Error of Measurement (SEM) of 3° [160]. Five studies investigated the intra-rater reliability for extension [5, 32, 50, 129, 160], showing fair to excellent reliability. One study calculated a SEM of 2° [160].

Four studies reported the intra-rater ICC's for pronation and supination, all showing excellent reliability [5, 55, 60, 160]. Two studies used the SEM, showing 3° and 7° respectively for pronation and 3° and 4° for supination [55, 160]. One study investigated the ICC in measurements of the carrying angle, showing excellent reliability [32].

Interrater reliability

Six studies investigated the interrater reliability for flexion, showing fair to excellent reliability. Results for expert and non-expert raters [19] and passive and active measurements [160] were similar. Two studies reported a SEM of 2° and 5° respectively [90, 160].

Five studies tested the inter-rater reliability in extension, showing fair to excellent reliability [5, 19, 115, 129, 160]. Results for expert and non-expert raters [19] and passive and active measurements [160] were similar. Zwerus et al. reported a SEM of 2° [160].

Five studies investigated the interrater reliability of pronation and supination [5, 35, 55, 56, 160], all showing excellent reliability. Results for injured and non-injured subjects [35] and passive and active measurements [160] were similar. Zwerus et al. reported a SEM of 3° for both pronation and supination [160].

		Intra-rater reliability								Inter-rater reliability							
Study		Flexion		Extension		Pronation		Supination		Flexion		Extension		Pronation		Supination	
		ICC	SEM	ICC	SEM	ICC	SEM	ICC	SEM	ICC	SEM	ICC	SEM	ICC	SEM	ICC	SEM
Armstrong 1998	Expert (n=3)	0.55-0.98	-	0.45-0.98	-	0.96-0.99	-	0.96-0.99	-	0.58-0.62	-	0.58-0.87	-	0.83-0.86	-	0.91-0.93	-
	Non-expert (n=2)	0.59-0.79	-	0.97-0.98	-	0.97	-	0.97-0.98	-								
Blonna 2012	Expert (n=3)	-	-	-	-	-	-	-	-	0.94-0.98	-	0.94-0.98			-	-	-
	Non-expert (n=1)	-	-	-	-	-	-	-	-	0.81-0.86	-	0.76-0.78			-	-	-
Chapleau 2011		0.95	-	0.97	-	-	-	-	-	-	-	-	-	-	-	-	-
Cimatti 2013	Injured	-	-	-	-	-	-	-	-	-	-	-	-	0.94	-	0.97	-
	Non-injured	-	-	-	-	-	-	-	-	-	-	-	-	0.92	-	0.95	-
Fieseler 2015		0.79-0.96	-	0.80-0.88	-	-	-	-	-	-	-	-	-	-	-	-	-
Flowers 2001		-	-	-	-	-	7.0	-	3.7	-	-	-	-	0.79	-	0.95	-
Gajdosik 2001		-	-	-	-	0.81-0.97	-	0.81-0.97	-	-	-	-	-	-	-	-	-
Goodwin 1992		0.61-0.92	-	-	-	-	-	-	-	0.56-0.91	-	-	-	-	-	-	-
Low 1976		-	-	-	-	-	-	-	-	-	5	-	-	-	-	-	-
Petherick 1987		-	-	-	-	-	-	-	-	0.53	-	0.53	-	-	-	-	-
Rothstein 1983		0.94-0.97	-	0.86-0.99	-	-	-	-	-	0.89-0.96	-	0.93-0.96	-	-	-	-	-
Zwerus 2017	Active	0.76	3	0.92	2	0.90	3	0.91	3	0.86	2	0.89	1	0.92	3	0.87	3
	Passive	0.74	3	0.95	2	0.86	3	0.90	3	0.79	2	0.85	2	0.91	3	0.82	3

Table III – Inter and intra-rater reliability of included studies

Discussion

The reliability and validity of the UG in measurements of the elbow was systematically examined. Based on 12 included studies, the overall reliability of the UG ranged amongst studies, from poor to excellent. There was no clear difference between intra- and inter rater reliability. The most striking outlier included deviating measurements of one expert-rater for inter- and intrarater reliability for flexion and extension in the study by Armstrong et al.[5] without providing a clear explanation.

The reliability for flexion, extension, pronation and supination was similar. The hypothesis that the reliability of the UG is higher in the hands of an expert examiner seems partially true[32]. Amstrong et al. did not find a difference in intrarater reliability in expert versus non-expert examiners, but they did give all examiners specific directions about arm positions and goniometric alignment [5]. In the study from Blonna et al., the examiners were free to use any bony landmarks they preferred [19]. They found a lower interrater reliability in non-expert examiners compared to expert examiners. This suggests that the reliability of the UG in non-expert examiners can be easily increased by clear instructions on goniometric alignment.

Previous literature stated that the reliability of the goniometer is higher when the same bony landmarks are used [52]. In this systematic review the studies using bony landmarks may not show a higher reliability. It is important to mention though that three out of four studies using the bony landmarks were of moderate quality.

Only one of the included studies in our systematic literature review investigated the validity of the goniometer [32]. They used radiography as reference test for goniometric measurements and found a potential maximum error of 11.5%. When precise values of ROM of the elbow are needed, they advised radiographic measurements.

Several previous studies investigated the reliability of the UG in measurements of joints other than the elbow. For example, in a study by Brosseau et al. an excellent intrarater reliability of the UG for knee flexion was found [25]. They also stated that a difference of more than 5.5 degrees in knee flexion is necessary to determine progression/change in the range of motion. Kim et al. investigated the reliability of the UG in hip and shoulder measurement and found high test-retest reliability even in unskilled examiners [78].

In this digital era it is important to realize that a lot of research has been performed comparing the UG with other devises and methods such as; an internet goniometer, a digital goniometer and VDO clip based goniometry [20, 48, 71, 98, 100, 108, 113, 155]. To maximize homogeneity this review focused on the UG. A future systematic review can be performed including and comparing these devises. It might be interesting to compare these devises and measuring methods with radiographic measurements to objectify their validity.

In all systematic reviews, there is a risk of overlooking papers. To minimize this risk an extensive search with sensitive search criteria and synonyms was performed, in collaboration with an experienced librarian. Also the included papers were scanned for other suitable studies.

Another limitation is the diversity and heterogeneity of the included articles. To avoid this clinical heterogeneity strict inclusion criteria were applied. Nonetheless there was a high diversity in study methods, such as blinding or not blinding the examiners from their own or other measurements. There was a high difference in interval of measurements, which can influence the outcomes. Furthermore, four articles did not clarify whether the examiners were expert or non-expert examiners, which makes it more difficult to interpret the outcomes. Finally only five studies were of high quality; the other seven studies were of moderate of low quality. The strength of this article is that it gives a clear overview of the research performed and their outcomes.

Conclusions drawn from this literature review are also limited because of the use of ICC's to assess reliability. It would be favourable to use a different approach to assess the agreement between two quantitative methods of measurement, because it possibly draws a misleading conclusion and is hard to transfer to an individual patient.

The ICC is in the general literature defined as a ratio of variance of interest over total variance (composed of variance of interest and error variance). In reliability studies for range of motion measurement, the variance among patients is often considered as the variance of interest[128]. Because the ICC uses variance between subjects' ROM measurements to calculate reliability, a large variation between subjects will lead to a higher ICC, even though the measurement error is similar[135]. This could possibly draw a misleading conclusion of good reliability. For example, this could have been the case with the higher ICC's for measurements in injured subjects compared to non-injured subjects and/or the higher ICC's in non-experts compared to expert examiners[5, 35]. Furthermore, ICC's are not presented as metric units and can therefore not be directly applied on an individual.

There are several ways to avoid the aforementioned problems induced by the use of ICC. Some authors already made efforts to use other ways to assess the reliability such as the SD, SEM and SDD [55, 90, 160]. Contribution of variance caused by subjects ($Var_{subject}$), occasion ($Var_{occasion}$) or measurement error (Var_{error}) can be determined using variance components analysis, in order to calculate the Standard Error of Measurement (SEM) and the Smallest Detectable Difference (SDD). SEM can be calculated using the following formula: $SEM = \sqrt{(Var_{occasion} + Var_{error})}$ and SDD using the following formula: $SDD = \sqrt{2} * 1.96 * \sqrt{(Var_{occasion} + Var_{error})}$. These measurement focus on the variance of different sources of error instead of the ratio of variances (ICC) and are presented in the metric unit of the measurements (degrees, in our case), which makes it easier to interpret for the use in clinical practice[128, 135]. Bland and Altman (B&A) proposed an alternative analysis, based on the mean difference and limits of agreement[18]. B&A plot analysis evaluates a bias between mean differences and estimates an agreement interval in which 95% of the differences between two measurements fall. Based on this plot (presented in a certain unit or percentage), the clinician can decide whether the limits are acceptable or not. Therefore we suggest the use SEM/SDD supplemented with B&A analysis for future research on the reliability of goniometric measurements.

Conclusion

Twelve studies reported on the reliability of the UG in measurements of the elbow were included. Overall, the studies showed at least a fair intra- and interrater reliability of the UG. The reliability of the UG in non-expert examiners can be increased by clear instructions on goniometric alignment. For future research, it would be favourable to use another statistical approach to substitute or supplement to ICC's.

Chapter 3

Validity and reliability of elbow range of motion measurements using digital photographs, movies and a goniometry smartphone application

Renée Keijsers MD[1], Elisa L. Zwerus MD[2], Dagmar R.M. van Lith MSc[3], Koen L.M. Koenraadt PhD[4], Pjotr Goossens MSc[4], Bertram The MD PhD[1], Michel P.J. van den Bekerom MD[5], Prof. Denise Eygendaal MD PhD[1,2]

1 Department of Orthopaedic Surgery, Amphia Hospital, Breda, the Netherlands
2 Department of Orthopaedic Surgery, AMC, Amsterdam, the Netherlands
[3]Fysiotherapy Erik Zuur, Raamsdonksveer, the Netherlands
[4]Foundation for Orthopaedic Research, Care and Education, Amphia Hospital, Breda, the Netherlands
[5]Shoulder and Elbow Unit, Department of Orthopedic Surgery, Onze Lieve Vrouwe Gasthuis, Amsterdam, the Netherlands

J Sports Med. 2018 Nov 29;2018:7906875. doi: 10.1155/2018/7906875.

Abstract

Introduction. Range of motion (ROM) is closely monitored before and after surgery for stiff elbow and during rehabilitation. Measurements in the home environment may be helpful to increase involvement and adherence of the patient. Therefore, our objective is to investigate the validity and inter- and intra-observer reliability of 3 alternative methods to assess the ROM by the patient in a home-based situation, in comparison to the universal goniometer (UG). We hypothesize that all 3 alternative methods will be valid alternatives and show a level of reliability equivalent to UG.

Methods. Goniometric measurements of elbow flexion, extension, pronation and supination using photography, movie and a smartphone application were obtained. The validity of these measurement methods were compared to UG. The inter-observer and intra-observer reliability was calculated for all measurement methods.

Results. Photography and movie based goniometry of the elbow showed good validity in flexion and extension. The inter-observer and intra-observer reliability where found to be good to excellent for photo and movie, but moderate to poor for UG and the smartphone application.

Conclusions. Photo or movie based goniometry seems to be a useful option for initial and follow-up measurement of the elbow ROM, both in the outpatient clinic and in a home environment. Based on our study, the smartphone application we used is not recommended.

Introduction

Reliable measurement of the range of motion (ROM) of the elbow is important for both the initial assessment and at follow up, to assess the results of surgery or to monitor rehabilitation. Using one reliable and reproducible measurement method between healthcare professionals and the patient himself, is beneficial to monitor the effect of different interventions.

In general, the ROM is assessed by an examiner by visual estimation or the use of an universal goniometer (UG) [57]. Amongst the possible alternative measurement methods are for example photography, movie or a smartphone based application [20, 48, 100, 108, 113, 155]. Photography and movie based measurements have several advantages compared to UG. For example, it provides a permanent image, which can be used to visually demonstrate the improvement to the patient, possibly helping to improve patient adherence. This relationship has been proved in other medical fields, for example in measurements of blood pressure in the home environment. In literature a rise of almost 10% in medicine compliance and significant blood-pressure reduction has been described [10, 24]. However, for those measurement methods a transfer from the camera or mobile phone to a computer with measurement software is necessary. Also, it requires an assisting person to take the picture or movie.

Smartphone based applications, based either on photography or an accelerometer, may be less time consuming and easy to use in a home environment. Measuring at home gives the patient the possibility to take more responsibility for his rehabilitation and update the health care provider between appointments.

Previous studies showed that both visual estimation and UG measurement have a good to excellent inter- and intra-observer reliability [5, 19, 32, 50, 52, 56, 57, 68, 74, 115, 129, 153, 160]. In literature on several other joints, excellent reliability for photography and smartphone apps was observed. Studies include photography of the elbow[20, 52] or knee[108] and smartphone apps based on an accelerometer principle for the knee[48, 100, 113] or shoulder[155].

For the elbow specifically, it is unclear which of the available goniometric measurement methods is the most reliable to measure flexion, extension, pronation and supination. Therefore, the objective of the current study is to investigate the validity and reliability of photography, movie and smartphone application based goniometry compared to UG. We hypothesize that all 3 alternative methods to measure elbow ROM will show a similar level of accuracy and inter- and intra-observer reliability, equivalent to UG.

Materials & methods

Study design

In this study measurements of the ROM of the elbow were collected using the UG, a smartphone application, photography and movies. The pictures and the movies itself were made by independent volunteering 'photographers', in general a family member who accompanies the participant. UG measurements were conducted by two health care professionals: a resident in the orthopedic department and a physiotherapist. The smartphone application measurements were done by the patient, with the aforementioned two health care professionals reporting the outcomes. Ethical approval was waived by the local ethical committee.

Study population

Subjects with or without elbow pathology, accompanied by a volunteer able to handle a digital camera ('photographer'), were included. Both must be 18 years or older, have sufficient knowledge of the Dutch language and physically and cognitively able to perform the proceedings in the measurement protocol. Subjects were recruited in a general hospital, a sports- and performance center and physiotherapy clinic. These locations and flexible inclusion criteria were chosen to ensure adequate representation of a population with and without elbow complaints.

Study procedure

Demographic data on height, weight, age, gender and hand dominance were collected for each subject. In all subjects the active ROM of the elbow (dominant side) was measured three times with UG and three times by an application on a Smartphone, by both observers independently. The order of the method of measurement by the two observers (UG vs application and examiner 1 vs examiner 2) was randomized by means of block randomization (blocks of 4). The photos were taken twice and the movie once by the 'photographer' in stated order, after each block of UG and Smartphone App sessions.

Subjects were instructed to carry out the four positions of the ROM: maximum flexion, extension and functional forearm rotation in pro- and supination. Attention must be payed to the difference between functional forearm rotation and pronation and supination. The functional forearm rotation measures the motion of forearm rotation in the two radioulnar joints (proximal and distal), combined with carpal rotation. Pronation and supination measures only the motion of the two radioulnair joints and is therefore a few degrees smaller. A study by Cimatti et al. showed that both methods could be used in clinical practice with excellent reliability [35]. In this study it is decided to use the functional forearm rotation because it is easier to implement for laymen. This means that in our results supination stands for forearm rotation in supination direction and pronation for forearm rotation in pronation direction.

<u>Universal goniometer</u>

Two observers measured all subjects' ROM three times independently with an UG. Between measurements of the two observers in one subject, a minimal interval of 5 minutes was applied. A stainless-steel goniometer was used and measurements were blinded for the observer by reversing the goniometer. Measurements were recorded with accuracy of 1 degree. A predefined protocol was used by both observers, based on recommendations in previous literature by using bony landmarks [27, 36, 50, 52, 117, 137, 140, 150, 154, 160].

For flexion and extension measurements of the elbow, the shoulder was held in 90 degrees of forward flexion with the forearm maximally supinated. The acromion and radial styloid process were landmarks for the goniometers' arms and the lateral epicondyle as the center of rotation. Supination and pronation were measured with a neutral position of the shoulder (0° shoulder abduction) and 90° of elbow flexion and a pencil placed over the distal palmar groove of the hand. The center of rotation for pronation and supination was over the head of the third metacarpal and the goniometers' arms were placed parallel to the humeral midline and parallel to the pencil.

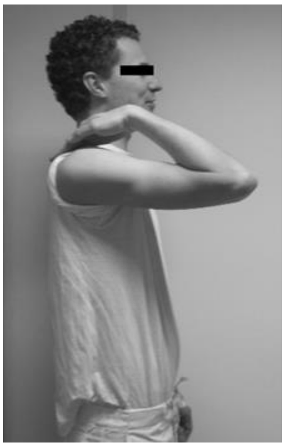


Figure 1A – Elbow flexion

Figure 1B – Elbow extension

Figure 1C – Functional forearm pronation

Figure 1D – Functional forearm supination

Photography

The photographers were instructed how to take the photos by a comprehensive and simplified protocol with sample pictures (Figure 1 A-D). The positions and motions were standardized as for the UG measurements described in the previous paragraph. The 'photographer' takes two series photos of the subject with a minimal interval of 20 minutes using a digital camera. In total 8 pictures were taken by each photographer. Elbow ROM on pictures was measured two times by both observers separately with a minimal interval of one day, using Kinovea software (Version 0.8.15, open source project, www.kinovea.org).

Movie

The movie was made by the photographer using the same protocol and device as for the photo as described in the previous paragraph. Subjects were instructed to slowly (in 5 seconds) carry out the movements from maximum flexion to maximum extension (movie 1) and from maximum supination to maximum pronation (movie 2). In total two movies, one for flexion-extension and one for pronation-supination movement, were taken by each photographer. Elbow ROM on the same movie was measured twice with a minimal interval of one day by each of two aforementioned observers separately using Kinovea software.

Smartphone application

The Joint Goniometry application (version 2.1, Diomidis Papas via App Store) for smartphones was used in simple mode for the elbow ROM measurements. This app is based on the principle of an accelerometer, comparable to other accelerometer based smartphone applications available in the App Store and Google Play. All 4 movements, as mentioned in de previous paragraphs, were measured three times by the two aforementioned observers. The previously mentioned landmarks were used as for the UG. When the subjects arm was held in the right position, the smartphone was placed with the middle of the bottom on the center of rotation and aligned to the proximal arm. The correct position was confirmed by a tap on the screen, followed by alignment on the distal arm and again confirmed by another tap. Measurements were blinded using a non-transparent elastic band on the screen.

Data and statistical analysis

In the study preparation phase, the sample size was calculated. Based on a significance of 0.05 (alpha) and power of 0.20 (beta), assuming a moderate correlation for our four measurement modalities, at least 18 participants were required.

All data were analyzed using SPSS version 22 (Armonk, NY, USA: IBM Corp) and Medcalc (version 16.1). A p-value of <0.05 was considered statistically significant. Data was checked manually for outliers in distribution. Subject characteristics are presented using descriptive statistics and 95% confidence intervals (CI).

Photography, movie and smartphone application based goniometry were individually compared to the UG measurements to analyze validity. The mean of the measurements in all three methods (photo, movie and app) of both observers was compared to the mean of the measurements with the UG. The agreement between the alternative measurement methods and the UG was calculated using the intra-class correlation coefficient (ICC). Because the ICC uses variance between subjects' ROM measurements to calculate reliability, a large variation between subjects will lead to a higher ICC. This could possibly draw a misleading conclusion of good reliability [128, 135]. Therefore we decided in our study to provide the mean difference (Δ) and accompanying 95% confidence intervals (CI) as well. To calculate these values, the mean of all three goniometer measurements (for maximum flexion, extension, pro- and supination separately) was compared to the mean of all three measurements by photo, movie or app.

For the inter-observer reliability, the same photo (photo 2) and movie were measured by both observers. For the smartphone application and UG the second measurement of both observers were compared. In the same way, the mean difference (Δ) and 95% CI were determined; the means of the measurements of observer 1 were compared to the means of observer 2. The intra-observer reliability was determined based on the measurements of the first observer (resident in the orthopedic department). The measurements of photo 1 were compared to photo 2. For the smartphone application and the UG the measurements of all three moments were compared. For the movie two measurement moments of the same movie were compared. Again, Δ and SD were determined. The inter-observer and intra-observer reliability was calculated using ICC.

For both validity and reliability analysis, the ICC's were calculated using a two-way random effects model where both people effects and measures effects are random. ICC between 0.75 and 1.00 indicates excellent reliability, between 0.60 and 0.74 good, between 0.40 and 0.59 moderate and ICC of ≤0.40 indicates a poor reliability[54]. Bland-Altman plots defining the limits of agreement (LOA) were used to determine whether a good correlation between two measurement methods also means that there is a good agreement between two methods[18]. A t-test was subsequently conducted to check for systematic errors. In addition, the linear regression was examined to check for proportional errors.

Results

Subject demographics

The study included 40 subjects (21 males and 19 females), each accompanied by an inexperienced 'photographer'. One subject had an elbow disorder without a functional disability. The mean age was 48 years (95% CI 43-54), mean height was 175 centimeters (95% CI 172-177) and mean weight 83 kilograms (95% CI 78-89). Four subjects (10%) were left-handed and 36 right-handed.

Validity of measurement methods

For flexion and extension, both photography and movie based measurements show a good to excellent correlation with UG. In pronation and supination measurement using photography and movie showed a moderate correlation with UG. The correlation between photography and movie measurements was good. The smartphone application correlated good with the UG in pronation and supination. Poor correlation for the smartphone application was shown for extension measurement, while flexion showed a moderate correlation. A proportional error was observed for extension in both photo and movie. This means that with increasing angles, the difference in angle between photo and movie with UG increased. The validity for the photography-, movie- and smartphone application based measurement methods compared to UG, the Δ and 95% confidence intervals (CI) are reported in table 1.

UG vs.		Photography	Movie	Smartphone application
Flexion	ICC (95% CI) Δ (95% CI) (°)	0.71 (0.51-0.83) 0 (-1.9 to 1.9)	0.63 (0.41-0.79) 0 (-1.9 to 1.9)	0.57 (0.32-0.75) 0 (-2.2 to 2.2)
Extension	ICC (95% CI) Δ (95% CI) (°)	0.76 (0.58-0.87) 1 (-0.2 to 2.2)	0.78 (0.63-0.88) 0 (-1.2 to 1.2)	0.28 (-0.05-0.55) 5 (3.8 to 6.2)
Pronation	ICC (95% CI) Δ (95% CI) (°)	0.44 (0.15-0.66) 4 (1.5 to 6.5)	0.45 (0.17-0.67) 2 (-0.5 to 4.5)	0.67 (0.47-0.82) 1 (-0.5 to 2.5)
Supination	ICC (95% CI) Δ (95% CI) (°)	0.50 (0.23-0.70) 2 (-0.8 to 4.8)	0.47 (0.18-0.68) 1 (-2.4 to 4.4)	0.61 (0.37-0.77) 1 (-1.5 to 3.5)

Table 1 – Validity of different measurement methods compared to UG measurements (ICC) and mean difference (Δ) and 95% confidence intervals (CI) (in degrees).

Inter-observer and intra-observer reliability analysis

The inter-observer reliability was excellent for photography and movie based measurements (table 2). Results for the smartphone application and UG were moderate to good. The mean differences between observers' measurements, are in all cases less than 5 degrees, however the accompanying 95% CI show a very wide range for UG and the smartphone application.

		UG	Photography	Movie	Smartphone application
Flexion	ICC (95% CI) Δ (95% CI) (°)	0.41 (0.07-0.65) 5 (2.8 to 7.2)	0.83 (0.65-0.92) 1 (0.1 to 1.9)	0.86 (0.75-0.92) 1 (0.1 to 1.9)	0.66 (0.45-0.81) 2 (-0.5 to 4.5)
Extension	ICC (95% CI) Δ (95% CI) (°)	0.65 (0.43-0.80) 1 (-0.5 to 2.5)	0.93 (0.88-0.96) 0 (-0.9 to 0.9)	0.88 (0.77-0.93) 1 (-0.2 to 2.2)	0.56 (0.31-0.74) 2 (-0.2 to 4.2)
Pronation	ICC (95% CI) Δ (95% CI) (°)	0.40 (0.11-0.63) 2 (-1.4 to 5.4)	0.90 (0.76-0.95) 3 (1.5 to 4.5)	0.82 (0.56-0.91) 3 (1.5 to 4.5)	0.55 (0.29-0.73) 3 (-1.3 to 7.3)

Supination	ICC (95% CI) Δ (95% CI) (°)	0.71 (0.51-0.83) 0 (-2.2 to 2.2)	0.89 (0.77-0.94) 2 (0.5 to 3.5)	0.96 (0.93-0.98) 1 (0.1 to1.9)	0.48 (0.20-0.67) 1 (-2.7 to 4.7)

Table 2 – Inter-observer reliability (ICC) and mean difference (Δ) and 95% CI (in degrees) of UG, photography, movie and smartphone application measurements.

Intra-observer reliability was good to excellent for photography based measurements and excellent for movie (table 3). The intra-observer reliability was poor to moderate for the smartphone application and moderate to excellent for UG. The mean differences between to measurements of the same observer are under 5 degrees, apart from the pronation measurement using the smartphone.

		UG	Photography	Movie	Smartphone application
Flexion	ICC (95% CI) Δ (95% CI) (°)	0.50 (0.31-0.67) 3 (2.1 to 3.9)	0.87 (0.81-0.92) 1 (0.7 to 1.3)	0.94(0.89-0.97) 1 (0.7 to 1.3)	0.60 (0.36-0.76) 4 (3.4 to 4.6)
Extension	ICC (95% CI) Δ (95% CI) (°)	0.84 (0.75-0.91) 2 (1.7 to 2.3)	0.82 (0.73-0.88) 1 (0.7 to 1.3)	0.96(0.93-0.98) 1 (0.7 to 1.3)	0.45 (0.16-0.66) 3 (2.4 to 3.6)
Pronation	ICC (95% CI) Δ (95% CI) (°)	0.71 (0.57-0.81) 3 (2.4 to 3.6)	0.72 (0.59-0.81) 3 (2.1 to 3.9)	0.94(0.98-0.97) 2 (1.7 to 2.3)	0.58 (0.33-0.75) 6 (3.2 to 8.8)
Supination	ICC (95% CI) Δ (95% CI) (°)	0.47 (0.28-0.65) 4 (2.1 to 5.9)	0.71 (0.57-0.81) 3 (2.4 to 3.6)	0.95(0.92-0.98) 2 (1.7 to 2.3)	0.31 (0.02-0.56) 5 (3.1 to 6.9)

Table 3 – Intra-observer reliability (ICC) and mean difference (Δ) and 95% CI of UG, photography, movie and smartphone application measurements

Discussion

The current study reported validity, inter-observer and intra-observer reliability for universal goniometry compared to 3 alternative measurement methods for elbow goniometry including photography, movie and a smartphone application. Validity appeared to be dependent on which elbow motion was measured. Photography and movie based goniometry showed better validity in flexion and extension, whereas the smartphone application showed better validity for pronation and supination. With respect to the reliability, inter-observer and intra-observer reliability were found to be good to excellent for photo and movie, but were predominantly poor to moderate for UG and the smartphone application. This means that in our study the variance in measurements amongst and within the observers is smaller for photo and movie compared to UG and the smartphone application.

In our study a systematic (proportional) error underestimating the extension measurement was observed by both photo and movie when compared to UG. Therefore, the results of extension from photo and movie are not interchangeable with UG. These findings are in line with previous literature on elbow and knee goniometry [20, 48, 52]. It is questionable if this error is caused by the photo, movie or UG measurement. Difficulties identifying the rotation center landmark has been designated as source for an error in the extension using photography or UG. Hence, it seems that in literature the UG is underestimating the extension angle [20, 52, 108].

With respect to the reliability of the UG measurements, our study results are only partially in line with previous literature. Literature on inter-observer and intra-observer reliability shows ICC values within a wide range, from 0.45-0.99, yet most ICC's were over 0.70 [5, 32, 50, 56, 115, 129, 160]. In our study, the inter-observer reliability of UG were moderate to good, ranging from 0.40 to 0.71 and the intra-observer reliability was moderate to excellent, ranging from 0.47 to 0.84. The wide range for reliability in both the literature and our study could be explained by the fact that the observers only had a few years' experience.

The reliability of photography in our study is in line with previous studies, however for the smartphone application our study demonstrated lower reliability. In literature, for both photography and smartphone apps excellent reliability was observed for several joints. Studies include photography of the elbow[20, 52] or knee[108] and smartphone apps based on an accelerometer principle for the knee[48, 100, 113] or shoulder[155]. All studies showed that photography or a smartphone application offer better reliability and are less dependent on the observers' experience compared to UG. A possible explanation for the disappointing results for the smartphone application in our study is the use by laymen. When tapping the screen, the application sometimes faltered and deviating results were not always recognized by the subjects.

We did not find literature using movie based goniometry. The excellent ICC's we found for the movie could be an overestimation, because two observations of measurements by each observer were based on a single movie.

Consideration should be given to the fact that UG might not be the most reliable method for elbow ROM measurement, especially in inexperienced examiners, as shown both in the literature and by the current study [19, 52]. Also, goniometry is used on a moving subject, unlike photography and movie, where measurements are carried out on a still image. Furthermore, measurement of functional forearm rotation (thus including carpal rotation) and pronation and supination are frequently placed under a common denominator. However, this accounts for all types of measurement methods we used.

This study is not without limitations. Subjects under 18 years old were excluded because of legal issues in younger patients. Moreover, in our study sample of 40 participants no patients with functional disabilities were included. Our results may not be automatically generalized for a population with elbow pathology without additional research. However, previous literature comparing the reliability of goniometric elbow measurements of pronation and supination show good inter- and intra-rater reliability for non-injured and even better for injured subjects [35, 74].

To verify correctness of measurements, measurements took place on our location, still simulating the home environment. It appeared that some participants required minimal adjustments to conduct the protocol correctly; in particular, during the imaging of the maximum supination to maximum pronation some of the participants forgot to keep the elbow against the body. For measurements in the home environment it is recommended to emphasize this in the protocol and for example practice the measurements with the patients the first time at the outpatient clinic or rehabilitation/physiotherapy center. It also might be illustrative to provide an accompanying instruction film when the measurements will actually take place in a home situation.

In order to obtain a measurement as reliable as possible, we recommend to use photography or movie for measurements both at the outpatient clinic and in the home environment. This allows the clinician to save the photo or movie and demonstrate the change (e.g. before and after intervention or follow-up) by showing sequential photos or movies to the patient. This provides the opportunity to increase patient engagement and adherence to rehabilitation therapy. Furthermore, between therapy sessions and for the long term follow-up the patient has to visit the clinic less frequently without losing important information on the patients' progress.

Conclusion

Based on this study, we recommend the use of photo or movie based goniometry for flexion and extension measurements of the elbow motion. These methods can be used in both the clinic and a home environment to increase the amount of follow-up moments and patient engagement during the rehabilitation process.

Part II

Physical examination of the elbow

Chapter 4

Physical examination of the elbow, what is the evidence? A systematic literature review

Elisa L. Zwerus, MD[1], Matthijs P. Somford, MD[2], François Maissan, MSc [3], Jelle Heisen, MSc[4], Denise Eygendaal, MD, PhD[5], Michel P.J. van den Bekerom, MD[1]

[1] Shoulder and elbow unit, Department of orthopaedic surgery, Onze Lieve Vrouwe Gasthuis, Amsterdam

[2] Department of orthopaedic surgery, Rijnstate Hospital, Arnhem

[3] Research group Lifestyle and Health, HU University of Applied Sciences, Utrecht

[4] Movamento Physiotherapy, Amsterdam

[5] Upper limb unit, Department of orthopaedic surgery, Amphia, Breda and University of Amsterdam

British Journal of Sports Medicine 2017 Mar 1. doi: 10.1136/bjsports-2016-096712.

Abstract

Objective: primary to provide an overview of diagnostic accuracy for clinical tests for common elbow (sport) injuries, secondary accompanied by reproducible instructions to perform these tests.

Design: a systematic literature review according to the PRISMA statement.

Data sources: a comprehensive literature search was performed in MEDLINE via PubMed and EMBASE.

Eligibility criteria: we included studies reporting diagnostic accuracy and a description on the performance for elbow tests, targeting the following conditions: distal biceps rupture, triceps rupture, posteromedial impingement, medial collateral ligament (MCL) insufficiency, posterolateral rotatory instability (PLRI), lateral epicondylitis, medial epicondylitis. After identifying the articles, the methodological quality was assessed using the QUADAS-2 checklist.

Results: our primary literature search yielded 1144 hits. After assessment 10 articles were included: 6 for distal biceps rupture, 1 for MCL insufficiency, 2 for PLRI, 1 for lateral epicondylitis. No articles were selected for triceps rupture, posteromedial impingement and medial epicondylitis. Quality assessment showed high or unclear risk of bias in 9 studies. We described 24 test procedures of which 14 tests contained data on diagnostic accuracy.

Conclusions: numerous clinical tests for the elbow were described in literature, seldom accompanied with data on diagnostic accuracy. None of the described tests can provide adequate certainty to rule in or rule out a disease based on sufficient diagnostic accuracy.

What are the new findings

1. Numerous clinical tests for the elbow are described in literature, seldom accompanied with data on diagnostic accuracy.

2. None of the described tests can provide adequate certainty to rule in or rule out a disease based on sufficient diagnostic accuracy.

3. All tests described in this review are performed in a clinical setting with a high pre-test probability, so diagnostic accuracy may be different in a general practice.

How might it impact on clinical practice in the near future?

1. Sufficient knowledge of underlying anatomical structures is important to interpret test results.

2. Descriptions for the execution of the tests provided in this review can be used to perform the tests in a similar manner.

3. Diagnostic accuracy derived from publications should be interpreted with understanding on methodology.

Introduction

The elbow is a complex joint and due to relative instability of the osseous articulations of the elbow, ligaments are required to provide elbow stability[105]. The relative instability makes the elbow prone to overuse injuries, mostly caused by sports (especially overhead throwing

athletes[73]) and work with repetitive elbow movements. Repetitive overhead throwing imparts high valgus and extension loads to the athlete's elbow, causing shear stress posteriorly and tensile stress medially with compression on the lateral side[53]. About 20% of the overuse injuries in the young athlete involve the elbow[72]. Acute elbow injuries are also common, making up approximately 15% of emergency department visits for upper-extremity musculoskeletal injuries[118].
After obtaining a medical history, physical examination is in most cases essential to make the diagnosis. A wide range of clinical tests is available and therefore it is challenging for an examiner to choose the right test to diagnose the patient's condition. If physical examination is reliable enough to rule in or rule out pathology, sometimes additional diagnostic imaging such as MRI could be omitted. By diagnosing through physical examination, less burden for the patient and less delay in diagnosis will be achieved in a cost-effective way. However, without adequate performance of the tests and knowledge on diagnostic accuracy, an examiner is not able to interpret physical examination of the elbow. Descriptions of elbow tests can be found in a few textbooks and various articles, although these were not recently updated and show variation in the execution of tests. An example is the valgus stress test, described with different angles of elbow flexion, ranging from 20° to 70°[112, 122, 132].

Therefore the objectives of the current study were (1) to investigate the diagnostic accuracy for the available diagnostic tests for specific elbow injuries and (2) to provide detailed instructions in order to effectuate accurate execution and interpretation by the examiner in a similar manner. We aim to apply these objectives for common acute and chronic elbow injuries and accompanying physical examination tests derived from literature and experts' opinions.

Materials and methods

This literature review was conducted according to the PRISMA statement[103].

Study criteria

Both prospective and retrospective observational study designs were included. Studies should describe and evaluate physical examination tests of the elbow to diagnose one or more of the following conditions: total distal biceps rupture, triceps rupture, posteromedial impingement, medial collateral ligament (MCL) insufficiency, posterolateral rotatory instability (PLRI), and lateral or medial epicondylitis. Selection of these injuries and diagnostic tests were based on expert opinions and previous literature[37-39, 67, 73, 81, 101, 136, 144, 152] (table 1). Physical examination is defined as a manually performed test on a living human adolescent or adult. The reference standard should be the most accurate diagnostic test (e.g. clinical test, diagnostic imaging or surgery) available for the targeted disease, according to available literature. However, since literature in this field can be ambiguous, expert opinions on the suitability of the reference standard were valued as decisive. Studies should report sufficient data to calculate different values of accuracy. The minimum number of patients was not pre-determined. There were no language restrictions, studies that were not published in English or Dutch were included if translation was possible.

Table 1 – Conditions and corresponding tests

Condition	*Test*
Total distal biceps rupture	Hook test Passive pronation supination test (PFP) Supination-Pronation test Biceps squeeze test / (Biceps) belly squeeze test Bicipital aponeurosis (BA) flex test Biceps crease interval (BCI) Biceps crease ratio (BCR)
Total triceps rupture	Triceps squeeze test / (Triceps) belly squeeze test
Posteromedial Impingement Syndrome	Arm bar test Posteromedial impingement test/valgus overload test
Medial collateral ligament (MCL) insufficiency	Moving valgus stress test Valgus stress test / Ligamentous instability test Milking manoeuvre
Posterolateral rotatory instability (PLRI)	Table-top relocation test Stand-up test/chair push-up test Push-up test

	Lateral pivot shift test (awake/under anaesthesia) / Posterolateral rotatory apprehension test Posterolateral drawer test
Lateral epicondylitis / Tennis elbow	Cozen's test Polk's test Maudsley's test/middle finger resistance/extension test Mill's test Kaplan sign/test Grip strength test (5%-8%-10% decrease)
Medial epicondylitis / Golfer's elbow	Epicondylitis medialis (shear) test/Golfer's elbow test Polk's test

Search strategy

A comprehensive electronic literature search for diagnostic studies was performed in collaboration with an experienced clinical librarian in MEDLINE via PubMed and EMBASE (earliest year until May 2016). The foundation of every search consisted keywords (MeSH and in all fields) elbow joint, elbow, diagnosis, diagnostic test, physical examination, measurement, sensitivity, specificity and accuracy. All searches were complemented by disease and test-specific keywords as mentioned in table 1. Reference lists of all included articles were manually checked for potential eligible citations.

Study selection

Two authors (MS and EZ) independently screened all titles and abstracts to identify relevant studies meeting the described study criteria. The authors were not blinded for author and affiliation names of these studies. Full text articles of potentially relevant studies were obtained and assessed. Any disagreement was solved by consensus by the help of a third author (MB).

Data extraction

Two authors (EZ and MS) extracted the following data from the included studies: first author, year of publication, study design, setting, number of patients, diagnosed pathology, prevalence, index test(s), reference test(s) and data to construct 2x2-tables for diagnostic accuracy properties.

Quality assessment

The quality of each included study was assessed by two authors (EZ and FM) using the revised version of the QUADAS, namely the QUADAS-2 checklist[156]. The checklist contains 4 domains: patient selection, index test, reference standard, and flow and timing. These domains are assessed for the risk of bias, and the first 3 domains are also assessed for applicability. Each item was scored “low”, “high” or “unclear”. Studies which scored “low” on all four domains received the overall judgement “low risk of bias and low concern regarding applicability”. If a study was judged "high" or "unclear" on one or more domains then overall judgement was “at risk of bias or concerns regarding applicability”.

Data analysis and synthesis

Data-analysis was performed by two independent authors (EZ and FM) using Microsoft Excel 2010 (Microsoft Corp. Washington, USA). Sensitivity/specificity, positive predictive value (PPV), negative predictive value (NPV) and likelihood ratios (LR) were calculated if possible. A positive likelihood ratio (LR+, i.e. post-test probability of disease presence) >10 and a negative likelihood ratio (LR−, i.e. post-test probability of disease absence) <0.1 were defined as large[62]. Confidence intervals for sensitivity and specificity[110] and for positive and negative likelihood ratios were calculated[142].

Results

Study selection

A total of 1142 articles were identified and manual reference checking identified 2 extra articles. After removal of duplicates, 1114 articles were available for screening on title. After exclusion of 1014 articles the abstracts of the remaining 100 articles were reviewed. Twenty-four full-text articles were selected and assessed. Ten articles were included into the final selection (figure 1).

Quality assessment

The QUADAS-2 checklist showed a low risk of bias and applicability concerns on all items in 1 study[131]. All other studies had 1 to 4 items scored as 'high' or 'unclear risk and applicability concerns' (table 2).

Table 2 – Risk of bias and applicability assessment based on QUADAS-2 checklist

Condition	**Study**	**Risk of bias**				**Applicability concerns**		
		Patient selection	Index test	Reference standard	Flow and	Patient selection	Index test	Reference standard
Total distal biceps rupture	Metzman 2015	Unclear	Low	Low	Low	High	Low	Low
	Devereaux 2013	Low	High	Low	Low	High	Low	Low
	ElMaraghy 2013	High	Low	Low	Low	Low	High	Low
	ElMaraghy 2008	Low	Low	Unclear	Low	High	Low	Low
	O'Driscoll 2007	High	Low	Low	High	High	High	Low
	Ruland 2005	Low	Low	Low	Low	Low	Low	Low
MCL insufficiency	O'Driscoll 2005	Low	Low	Low	Low	Low	Low	Low
PLRI	Arvind 2006	High	Low	Low	Low	High	Low	Low
	Regan 2006	High	Low	Low	Low	High	Low	Low
Lateral epicondylitis	Dorf 2007	Low	Low	Unclear	Low	High	Low	Low

Study characteristics

Ten articles were included: 6 (170 patients, 134 with disease) for total distal biceps rupture [41, 46, 47, 99, 111, 131], 1 (21 patients, 17 with disease) for MCL insufficiency[112], 2 (16 patients, all with disease) for PLRI[9, 126] and 1 (40 patients, all with disease, the contralateral arm served as control group) for lateral epicondylitis[43]. No articles were selected for total triceps rupture, posteromedial impingement and medial epicondylitis. Characteristics of studies are summarized in table 3.

Table 3 – Summary of study characteristics

Condition	**Study**	**Design**	**Setting**	**Prevalence**	**Reference test**	**Index test**
Total distal biceps rupture	Metzman 2015[99]	Case control	General hospital	5/5 (100%)	Surgery	Supination-pronation test
	Devereaux 2013[41]	Cohort	Teaching hospital	42/48 (87.5%)	Surgery (T+) or MRI (T-)	Hook test
						PFP
						BCI
	ElMaraghy 2013[46]	Case control	Orthopaedic clinic	7/17 (41.2%)	Surgery	BA flex test
	ElMaraghy 2008[47]	Cohort	University hospital	24/29 (82.8%)	Surgery (T+) or MRI (T-)	BCI
						BCR
	O'Driscoll 2007[111]	Case control	Tertiary clinic	33/45 (73.3%)	Surgery	Hook test
	Ruland 2005[131]	Cohort	Naval hospital	23/26 (88.5%)	Surgery (T+) or MRI (T-)	Biceps squeeze
MCL insufficiency	O'Driscoll 2005[112]	Cohort	Tertiary clinic	17/21 (81.0%)	Surgery	Moving valgus stress

						Valgus Stress (pain)
						Valgus stress (laxity)
PLRI	Arvind 2006[9]	Unclear	General hospital	8/8 (100%)	Pivot Shift test	Table top relocation test
	Regan 2006[126]	Cohort	Sports Medicine Centre	8/8 (100%)	Surgery	Stand-up/ Chair push-up test
						Push-up test
						Pivot shift (awake)
						Pivot shift (anaest.)
Lateral epicondylitis	Dorf 2007[43]	Case control	University Hospital	40/40 (100%)	Tenderness of ECRB, Mills test, Maudsley's test and, resisted wrist extension	Grip strength test 5%
						8%
						10%

Diagnostic accuracy of tests

In total 24 test procedures are described, for fourteen of these tests data on diagnostic accuracy were available. For total distal biceps rupture 7 different tests were described in 6 studies[41, 47, 99, 111, 131]. These tests showed sensitivity ranging from 81 to 100% and specificity ranging from not applicable to 100%. The hook test is performed by hooking a finger under the distal biceps tendon from the lateral side[41, 111]. The passive forearm pronation (PFP) test[41], supination-pronation test[99] and biceps squeeze test[131] investigate the function of the biceps tendon: in a normal functioning tendon pro-nation/supination results in a change of the muscle belly outline and squeezing the biceps results in supination. In the BA flex test, the examiner needs to identify the BA from the medial side. One of the included studies showed that in 59% of the 17 patients they followed (16 total biceps ruptures and one high grade partial rupture) the BA remained intact[46]. The biceps crease interval (BCI) test uses the distance between the antecubital fossa and the start of the biceps curve, with 6.0cm (2 standard deviations larger than in a control group) as a cut-off value[41, 47]. The biceps crease ratio (BCR) is an additional test that compares the biceps crease between the injured and non-injured arm[47]. A combination of the hook test, PFP and BCI altered the sensitivity, but dropped the specificity compared to separate tests[41]. Combining BCI and BCR resulted in similar diagnostic accuracy as BCR separately[47].

No studies were found describing a clinical diagnostic test for total triceps rupture. In textbooks though, the triceps squeeze test is described[38]. Based on the same mechanism used for the biceps squeeze test, the test is performed by squeezing the muscle belly of the triceps and conclusion of triceps rupture is made in case of absence of extension in the elbow.

The posteromedial impingement syndrome is caused by a mechanical constraint by bony or soft tissue in the posteromedial side of the elbow. The arm bar test[39] and valgus overload test (or posteromedial impingement test)[44] provoke entrapment of the tissue in the posterior (and medial) olecranon fossa. No diagnostic accuracy studies of these tests were found.

One study focussed on the diagnostic tests for MCL insufficiency[112]. The moving valgus stress test provokes the stress in a throwing movement, reproducing pain in the “shear angle” from 120° to 70°. Sensitivity for a MCL insufficiency was 75% and specificity 100%. In the valgus stress test the examiner holds the elbow in 70° flexion and applies valgus stress. For pain as an outcome, the test showed 65% sensitivity and 50% specificity. However, by utilizing laxity as outcome, the test had a disappointing sensitivity but perfect specificity of 100%. In literature the milking manoeuvre was described, but no evidence on diagnostic accuracy has been found in our search[33, 39].

For PLRI, caused by an injury of the lateral ulnar collateral ligament, 4 different tests were studied in 2 articles. All tests aim to reproduce instability and are assessed by the existence of apprehension or dislocation of the radial head combined with distinctive pain. The table-top relocation test[9], stand-up/chair push-up test[126] and push-up test[126] show similar capacities for a positive finding when the disease was presence with sensitivity from 88 to 100%. The pivot shift test is performed on a fully supinated and extended elbow followed by a combination of valgus stress and axial compression while flexing the elbow and is positive when the radial head dislocates around 40° flexion. The test shows sensitivity of only 38% in the awake patient, but 100% sensitivity when the test is performed under anesthesia[126].

Lateral epicondylitis is a disorder of the wrist- and finger extensors and specific diagnostic tests aim to stretch these muscles. Active movement tests include dorsal flexion of the wrist resisted by the examiner (Cozen's test[94]) or by holding a book (Polk test[119]) and dorsal extension of the middle finger against resistance (Maudsley's test[43]). Mills test is performed by passively bringing the hand to palmar flexion and hereby stretching the extensors[43]. None of these tests were supplemented by information on diagnostic accuracy. Only one test, the grip strength test, presented sensitivity and specificity data[43]. The decrease of grip strength was determined for a decrease of 5%, 8% and 10% using a hand-held dynamometer. Sensitivity ranged from 85% to 78% and specificity from 80% to 90%. A disadvantage of the grip strength test is that the examiner needs a special device to examine the patient.

Medial epicondylitis is a disorder of the wrist- and finger flexors and specific diagnostic tests aim to stretch these muscles. The epicondylitis medialis test or golfers elbow test[73] is performed by active palmar flexion of the hand without resistance and Polk's test[119] adds resistance by letting the patient hold a book. No studies on diagnostic accuracy for these tests were found.

Diagnostic accuracy of the 14 available studies is summarized in table 4. Descriptions for the performance of all 24 tests are presented in table 5.

Table 4 – Summary of study diagnostic accuracy of tests

Condition	Study	Index test	Diagnostic accuracy					
			Sens (%) (CI)	Spec (%) (CI)	PPV (%) (CI)	NPV (%) (CI)	LR+ (CI)	LR- (CI)
Total distal biceps rupture	Metzman 2015[99]	Supination-pronation test	100 (47.8 - 100)	N/A	100 (47,8 - 100)	N/A	N/A	N/A
	Devereaux 2013[41]	Hook test	81 (65.9 - 91.4)	100 (54.1 - 100)	100 (89.7 - 100)	42.9 (17.7 - 71.1)	N/A	0.19 (0.1 - 0.36)
		PFP	95.2 (83.8 - 99.4)	100 (54.1 - 100)	100 (91.2 - 100)	75 (34.9 - 96.8)	N/A	0.05 (0.01 - 0.18)
		BCI	88.1 (74.4 - 96)	50 (11.8-88.19)	92.5 (79.6 - 98.4)	37.5 (8.5 - 75.5)	1.76 (0.79 - 3.95)	0.24 (0.08 - 0.75)
		Combination hook + PFP + BCI	100 (91.59 -100)	50 (11.8-88.2)	93.3 (81.7-98.6)	100 (29.24 -100)	2.0 (0.90 - 4.45)	0

	ElMaraghy 2013[46]	BA flex test	100 (59 - 100)	90 (55.5 - 99.75)	87.5 (47.4 - 99.7)	100 (66.4 - 100)	10 (1.56 - 64.2)	0
	ElMaraghy 2008[47]	BCI	91.7 (73 - 99)	100 (47.8 - 100)	100 (84.6 - 100)	71.4 (29 - 96.3)	N/A	0.08 (0.02 - 0.3)
		BCR	95.8 (78.9 - 99.9)	80 (28.4 - 99.5)	95.8 (78.9 - 99.9)	80 (28.4 - 99.5)	4.79 (0.83 - 27.7)	0.05 (0.01 - 0.37)
		Combination BCI + BCR	95.8 (78.9 - 99.9)	80 (28.4 - 99.5)	95.8 (78.9 - 99.9)	80 (28.4 - 99.5)	4.79 (0.83 - 27.7)	0.05 (0.01 - 0.37)
	O'Driscoll 2007[111]	Hook test	100 (89.4 - 100)	100 (73.5 - 100)	100 (89.4 - 100)	100 (74.5 - 100)	N/A	0
	Ruland 2005[131]	Biceps squeeze	100 (85.2 - 100)	66.7 (9.4 - 99.2)	95.8 (78.9 - 99.9)	100 (15.8 - 100)	3 (0.6 - 14.9)	0
MCL insufficiency	O'Driscoll 2005[112]	Moving valgus stress	100 (80.5 - 100)	75 (19.4 - 99.4)	94.4 (72.7 - 99.86)	100 (29.2 - 100)	4 (0.7 - 21.8)	0

		Valgus Stress (pain)	64.7 (38.3 - 85.8)	50 (6.8 - 93.2)	84.6 (54.6 - 98.1)	25 (3.2 - 65.1)	1.29 (0.46 - 3.66)	0.71 (0.22 - 2.28)
		Valgus stress (laxity)	18.8 (4.1 - 45.7)	100 (38.8 - 100)	100 (29.2 - 100)	23.5 (6.8 - 49.9)	N/A	0.81 (0.6 - 1.03)
PLRI	Arvind 2006[9]	Table top relocation test	100 (63.1 - 100)	N/A	100 (63.1 - 100)	N/A	N/A	N/A
	Regan 2006[126]	Stand-up/ Chair push-up test	87.5 (47.4 - 99.7)	N/A	100 (59 - 100)	0 (0 - 97.5)	N/A	N/A
		Push-up test	87.5 (47.4 - 99.7)	N/A	100 (59 - 100)	0 (0 - 97.5)	N/A	N/A
		Pivot shift (awake)	37.5 (8.5 - 75.5)	N/A	100 (29.2 - 100)	0 (0 - 52.2)	N/A	N/A
		Pivot shift (anaest.)	100 (63.1 - 100)	N/A	100 (63.1 - 100)	N/A	N/A	N/A
Lateral epicondylitis	Dorf 2007[43]	Grip strength test 5%	83 (66.4 - 92.7)	80	N/A	N/A	4.2 (1.2 - 14.5)	0.21 (0.09 - 0.5)
		8%	80	85	N/A	N/A	5.33	0.24

			(58.4 - 91.9)	(64 - 94.8)			(1.8 - 15.5)	(0.1 - 0.6)
		10%	78 (59.1 - 88.2)	90 (59.6 - 98.2)	N/A	N/A	7.7 (1.2 - 49.7)	0.24 0.1 - 0.5)

T+ = index test positive *T- = index test negative*

PPV = Positive Predictive Value *NPV = Negative Predictive Value*

LR+ = Positive Likelihood Ratio *LR- = Negative Likelihood Ratio*

Table 5 - Description of specific diagnostic tests for the elbow

Total distal biceps rupture

Hook test[41, 111] (Sens 81-100% / Spec 100%)

Patient position	Seated, passive supination forearm, 90° elbow flexion
Examiner position	Index finger on antecubital fossa
Test	Hook index finger under intact biceps tendon from lateral side
Assessment	No cord-like structure to hook a finger indicates total distal biceps rupture, painful test indicates partial rupture

Passive forearm pronation test[41] (PFP) (Sens 95% / Spec 100%)

Patient position	Seated, 90° elbow flexion
Examiner position	Hand on m. biceps, fixate wrist
Test	Palpate m. biceps while pro-/supinating forearm passively
Assessment	No proximal excursion of biceps in supination and distal migration in pronation indicates total distal biceps rupture

Supination-pronation test[99] (Sens 100%)

Patient position	Standing, shoulders abducted 90°, elbows flexed 60-70°
Examiner position	Stand in front of patient, observe contour biceps

Test	Ask patient to actively supinate and pronate forearms by turning hands
Assessment	Lack of migration of the biceps muscle indicates total biceps rupture

Biceps squeeze test[131] *(Sens 100% / 67%)*

Patient position	Seated, forearm resting comfortably, 60°-80° elbow flexion
Examiner position	Hand on distal biceps tendon, other around muscle belly
Test	Pronate forearm slightly, squeeze both hands firmly
Assessment	No supination of the forearm indicates total distal biceps rupture

Bicipital aponeurosis (BA) Flex Test[46] *(Sens 100% / Spec 90%)*

Patient position	Seated, 0° elbow extension, active wrist flexion, forearm supinated
Examiner position	One hand on wrist, index finger on antecubital fossa
Test	Flex elbow passively to 75°, palpate medial/lateral/central parts of the antecubital fossa
Assessment	No sharp edge medially indicates BA rupture and could indicate total distal biceps rupture

Biceps crease interval[41, 47] *(BCI) (Sens 88-92% / Spec 50-100%)*

Patient position	Seated, 90° elbow flexion
Examiner position	Fixate wrist, index finger on antecubital fossa
Test	Passively extend the elbow, supinate forearm. Mark flexion crease in antecubital fossa. Mark start of biceps curve. Measure distance between marks
Assessment	Absolute BCI value >6cm indicates total distal biceps rupture

Biceps crease ratio[47] *(BCR) (Sens 96% / Spec 80%)*

Patient position	Seated, 90° elbow flexion
Examiner position	Fixate wrist, index finger on antecubital fossa
Test	Repeat steps of BCI test on contralateral arm, calculate ratio between BCI's in both arms

Assessment	BCR >1.2 indicates total distal biceps rupture

Total distal triceps rupture

Triceps squeeze test[38]

Patient position	Seated, forearm hanging comfortably over the back of a chair, 90° elbow flexion
Examiner position	Hand on distal triceps tendon, other around muscle belly
Test	Squeeze both hands firmly
Assessment	No extension of the elbow indicates total triceps rupture

Posteromedial impingement syndrome

Arm bar test[39]

Patient position	Standing, shoulder in full internal rotation (thumb pointing downwards) and 90° anteflexion, elbow extended, index finger resting on examiners' shoulder
Examiner position	Hand on distal humerus
Test	Apply pressure on distal humerus to fully extend patients' elbow
Assessment	Distinctive posteromedial pain indicates posteromedial impingement syndrome

Valgus overload test / posteromedial impingement test[44]

Patient position	Seated or standing, 20°-30° elbow flexion
Examiner position	Fixate upper arm and grasp wrist
Test	From starting position the examiner forcibly extends the elbow while applying valgus stress
Assessment	Distinctive posteromedial pain indicates posteromedial impingement syndrome

MCL insufficiency

Moving valgus stress test[112] *(Sens 100% / Spec 75%)*

Patient position	Seated, 90° shoulder abduction, maximum elbow flexion

Examiner position	Stabilize humerus and hold wrist
Test	Apply valgus stress until shoulder reaches maximum external rotation. Maintain valgus stress and quickly extend elbow to 30°
Assessment	Distinctive pain, max. between 120° and 70° flexion ("shear angle") indicates MCL insufficiency

Valgus stress test[112] *(Pain: Sens 65% / Spec 50%, Laxity: Sens 19% / Spec 100%)*

Patient position	Seated, 70° elbow flexion, supinated maximally
Examiner position	Stabilize humerus and hold forearm
Test	Apply valgus stress
Assessment	Distinctive pain or laxity (compare to other elbow) indicates MCL insufficiency

Milking manoeuvre[33, 39]

Patient position	Seated, shoulder 90° anteflexion, elbow >90° flexion, forearm supinated, fingers in fist, thumb pointing lateral
Examiner position	While maintaining patients position, stabilize humerus and grab thumb
Test	Apply downward and valgus stress on patients' thumb
Assessment	Distinctive pain indicates MCL insufficiency

Posterolateral Rotatory Instability (PLRI)

Table-top relocation test[9] *(Sens 100%)*

Patient position	Standing in front of a table, hand placed over lateral edge of the table, elbow pointing laterally and forearm supinated
Examiner position	Standing next to patient
Test	1) Ask patient to perform press-up with symptomatic arm; 2) Repeat with examinators' thumb on radial head; 3) Remove thumb with patient maintaining position in step 2
Assessment	1) Distinctive pain and positive apprehension at 40° elbow flexion; 2) Symptoms relieved;

3) Reproducing pain + positive apprehension

Stand-up test/chair push-up test[4, 126] *(Sens 88%)*

Patient position	Seated, both elbows 90° flexion, holding armrests with shoulder abduction and forearm supinated
Examiner position	Standing/sitting close to patient
Test	Ask patient to arise chair by pushing down
Assessment	Pain that slowly extends while patient rises indicates PLRI

Push-up test[9, 126] *(Sens 88%)*

Patient position	Lay with chest on the floor, elbows flexed at 90°, shoulders abducted, forearm supinated
Examiner position	Standing/sitting close to patient
Test	Ask patient to perform push-up
Assessment	Apprehension or radial head dislocation indicates PLRI

Lateral pivot shift test[9, 126] *(Awake: Sens 38%, Anaesthesia: Sens 100%)*

Patient position	Supine, shoulder anteflexion about 100° and full external rotation, forearm fully supinated, elbow maximally extended
Examiner position	Gasp patients' forearm and wrist
Test	Apply a combination of supination, valgus stress and axial compression to the elbow while flexing the elbow
Assessment	At approx. 40° flexion apprehension or dislocation of radial head (dimple in skin) indicates PLRI

Lateral Epicondylitis

Cozen's test[94]

Patient position	Seated, elbow extended, forearm maximal pronation, wrist radially abducted, hand in a fist
Examiner position	Stabilize elbow while palpating lateral epicondyle, other hand placed on dorsum of the hand

Test	Ask patient move the wrist to dorsal flexion and move the wrist towards palmar flexion
Assessment	Pain on the lateral epicondyle indicates lateral epicondylitis

Polk's test

lateral[119]

Patient position	Seated, elbow flexion about 100°, pronation of the forearm
Examiner position	Close/next to patient, carrying an object (e.g. a book) of approx. 2.5kg/5lb
Test	Ask patient to grab and lift the object
Assessment	Pain on the lateral epicondyle indicates lateral epicondylitis

Maudsley's test / Middle finger resistance test[43]

Patient position	Seated, extended elbow, forearm and palmar side of the hand on table
Examiner position	Stabilize forearm on table, finger on tip of digitus 3/middle finger
Test	Ask patient to lift middle finger while the examiner pushes the finger down to the table
Assessment	Pain on the lateral epicondyle indicates lateral epicondylitis

Mill's test[43]

Patient position	Seated, elbow extended, forearm pronated
Examiner position	Stabilize elbow while palpating lateral epicondyle, grab wrist
Test	Move the wrist passively in palmar flexion
Assessment	Pain on the lateral epicondyle indicates lateral epicondylitis

Grip strength test[43] *(5-8-10% decrease: Sens 83-80-78% / Spec 80-85-90%)*

Patient position	Seated, holding hand dynamometer with adducted shoulder, neutral rotation, forearm and wrist in neutral position
Examiner position	Seated next/close to patient

Test	Ask patient to squeeze the dynamometer as strong as possible (pain may occur) in 90° elbow flexion and secondly in full extension
Assessment	5%-8%-10% decrease in grip strength between flexion and extension indicates lateral epicondylitis

Medial epicondylitis

Epicondylitis medialis test / Golfer's elbow test

Patient position	Seated, elbow extended and fully supinated
Examiner position	Place one hand on the patients ventral side of the hand, stabilize the elbow with other hand
Test	Ask patient to move to hand to palmar flexion against your resistance
Assessment	Pain on the medial epicondyle indicates medial epicondylitis

Polk's test medial[119]

Patient position	Seated, elbow flexion about 100°, supination of the forearm
Examiner position	Close/next to patient, carrying an object (e.g. a book) of approx. 2.5kg/5lb
Test	Ask patient to grab and lift the object
Assessment	Pain on the medial epicondyle indicates medial epicondylitis

Discussion

The primary objective of the current study was to investigate the diagnostic accuracy for the available diagnostic tests for elbow injuries that mostly occur due to sports or work. Our secondary objective was to provide detailed instructions for these tests. In the current literature review, we focussed only on the diagnostic tests of the elbow and did not incorporate anamnesis and other findings in physical examination such as inspection, range of motion and palpation. Previous literature reviews of the elbow already gave an overview of the differential diagnosis of elbow pain based anatomic regions[73]. Other studies[67, 94, 102] and several books[37-39, 81, 101, 136, 144] elaborated on the history taking, physical examination and treatment of common elbow injuries. However, these resources are not up-to-date and did not conduct a systematic literature research. Valdes et al.[152] did a systematic review on provocative tests only, leaving out most of the common diagnostic elbow tests.

In literature, we found twenty-four test procedures described in such a way that the test is reproducible. For fourteen tests diagnostic accuracy was available. Interpretation of diagnostic accuracy results should be done with great caution. Sensitivity and specificity reflects the prevalence of a disease within a given population. PPV/NPV reflects the outcome of a test for an individual within a given population and is influenced by prevalence of disease[116]. For example; if there are few subjects with the disease, there is a reduced chance to diagnose the disease compared to when there are many subjects with the disease. As a result the PPV decreases. Because a lot of subjects in the example are symptomatic but do not have the investigated disease, the NPV increases. In the current review, most studied elbow conditions have a high prevalence in the studied population, in contrast to their prevalence in general orthopaedic practice. The higher prevalence compared to the general situation can result in a overestimated PPV in all studies. It is even possible, in studies with diseased people only, that PPV artificially rises to 100%[89]. In our review all PLRI tests[9, 126] and the supination-pronation test[99] showed an artificially high PPV based on diseased people only, resulting in even less transferable and generalizable diagnostic accuracy outcomes. Interpreting likelihood ratios from a clinical perspective, a LR+ >1 indicates that a positive test increases the post-test probability above the prevalence or pre-test probability. Thus, a LR+>1 means that the diagnostic test increases the probability that a condition is present[87]. However, none of the studies showed likelihood ratios that we defined as large. In most studies the CI of LR+ even crossed 1 and are therefore not significant. All grip strength tests for LE and the BA flex test for distal biceps rupture showed a statistically significant LR+[43, 46], however apart from the Grip strength test for 5% and 8% decrease the CI's were very wide. Therefore all tests contributed only little to the diagnostic certainty that the condition was present.

Amongst the included studies, two authors combined 2 or 3 diagnostic tests to obtain the diagnosis of distal biceps rupture. A combination of the hook test, PFP test and BCI showed an altered sensitivity, but resulted in a lower specificity compared to each individual tests[41]. Combining BCI and BCR on 17 patients showed the same diagnostic accuracy as for BCR[47].

The so-called "composite physical examination" (CPE) was evaluated in a few studies for the shoulder[69] and knee[45, 65, 84, 143, 146] joint, all showing slightly altered diagnostic accuracy, especially for sensitivity. Thus, a combination of tests with a positive result will in those cases increase the post-test probability. The limited studies on CPE for elbow injuries have not demonstrated an added value and needs further exploration.

In addition to the remarks on the interpretation of diagnostic accuracy statistics, important limitations in the quality of the included studies were observed. In general, small numbers of patients (ranging from 5 to 48) were included, explained by the fact that elbow conditions are relatively rare. The studies reporting on the supination-pronation test[99] had only 5 subjects and both studies describing tests for PLRI[9, 126] included only 8 subjects each. Conclusions based on these small sample sizes are neither transferable nor generalizable. Furthermore, due to the retrospective nature of most studies included in current systematic review, many patients were excluded because of an incomplete medical record. Only 2 out of 10 studies were conducted in a general hospital[9, 99], other hospitals were specialized clinics examining referred patients. Performing the tests in a specialized clinic could lead to selection bias, causing higher diagnostic accuracy in specialized clinics compared to the use in a general population. Only two studies concerning the biceps squeeze test[131] and tests for MCL injuries[112], met all the QUADAS-2 criteria and therefore seems to have low risk of bias. However, even in the abovementioned studies the number of symptomatic subjects without disease were still very low in comparison with the diseased subjects (3/26 and 4/21 respectively). Therefore PPV can be overrated.
Positive facts were that all studies used conventional reference tests for each condition such as MRI or surgery. Most studies described that reference test result was interpreted without knowledge of index test result.

Extensive knowledge of anatomy is important to perform the tests correctly; demonstrated by all described tests for total distal biceps rupture. These tests may show false negative results in patients with a chronic rupture, because of scar tissue or an intact BA that can easily be mistaken for the biceps tendon, leading to false negative results. O'Driscoll et al.[111] endorsed the importance of the recognition of anatomical structures by incorporating identification of the BA in their performance of the Hook test. The study showed excellent sensitivity and specificity identifying a total rupture, though in a much selected population. In all patients with a negative test, surgery showed a partial rupture. The BA flex test helps the examiner to identify the anatomy of the bicipital aponeurosis[46]. Moreover, the examiner should be aware that if the BA flex test shows a negative result (with NPV=100%), he can be fully certain that the BA is intact and may cause a false negative result for all bicep tests. Thus, performing a combination of tests can be essential to differentiate. A study evaluating the combination of the BA flex test and a distal biceps test (e.g. hook test) would be a valuable contribution to our current knowledge.

Initially, we aimed to perform a meta-analysis if more than one study was available on the same subject. Amongst the included studies, for the hook test and BCI two diagnostic studies were available for meta-analysis. However, all studies showed a high risk of bias, different study designs, a small sample size and diverse results with wide confidence intervals crossing the null hypothesis. Combining the studies would compound the errors and lead to an inappropriate conclusion[64, 88]. Therefore we decided to focus on a qualitative analysis. Furthermore, we designed the study to focus on 7 injuries, mostly sports related. We did not incorporate neurological problems such as cubital tunnel syndrome and elbow fractures. That decision was made because we believe that extensive amount of conditions would not favour the study.

Our study showed that further research is required in order to assist the physician in diagnosing elbow pathology. The need for further research applies in particular for the conditions for which we did not find a diagnostic study, e.g. triceps rupture (triceps squeeze test), posteromedial impingement syndrome (valgus overload test), lateral epicondylitis (Cozen's and Mills test) and medial epicondylitis (golfers elbow test). Previous studies and experiences in our daily practice showed us that using a combination of diagnostic tests could be favourable in order to obtain the diagnosis. However, based on the included studies performing elbow CPE, no conclusions for implementation can be drawn yet. Therefore we recommend a prospective cohort study about diagnostic accuracy of (a combination of) promising tests, reported following the Standards for Reporting Diagnostic accuracy studies (STARD) guidelines[22]. To prevent non-generalizable outcome on diagnostic accuracy, the population would preferably consist of a large symptomatic population, with about the same number of subjects with and without the condition as confirmed by an adequate reference test. Descriptions for the execution of the tests provided in the current review can be used to standardise procedures in future research.

Conclusion

Numerous clinical tests for the elbow are described in literature, seldom accompanied with data on diagnostic accuracy. The current literature review provides description of 24 tests for 7 conditions, whereof 14 tests with diagnostic accuracy. Specific diagnostic tests can be performed for the targeted elbow condition, taking into account the current available information on diagnostic accuracy and their interpretation. Based on our literature review, none of the described tests can provide adequate certainty to rule in or rule out a disease, based on sufficient diagnostic accuracy. Performing a combination of tests can be essential to differentiate. Moreover, adequate knowledge of the underlying anatomical structures is important to interpret test results correctly.

Chapter 5

Distal biceps tendon ruptures: diagnostic strategy through physical examination

Elisa L. Zwerus, MD[1,2], Derek F.P. van Deurzen, MD, PhD[3], Michel P.J. van den Bekerom, MD, PhD[3,4], Bertram The, MD, PhD[2], Denise Eygendaal, MD, PhD[1,2]

[1] Department of Orthopaedics and Sports Medicine, Erasmus University Medical Centre, Rotterdam, The Netherlands

[2] Department of Orthopaedic Surgery, Amphia Hospital, Breda, The Netherlands

[3] Shoulder and elbow unit, Department of Orthopaedic Surgery, OLVG, Amsterdam, The Netherlands

[4] Department of Human Movement Sciences, Vrije Universiteit, Amsterdam, The Netherlands

The American Journal of Sports Medicine 2022;50(14):3956–3962. DOI: 10.1177/03635465221129874

Abstract

Background: Distinguishing a complete from a partial distal biceps tendon rupture is essential, as a complete rupture may require repair on a short notice to restore function, whereas partial ruptures can be treated non-surgically in most cases. Reliability of physical examination is crucial in order to determine the right work-up and treatment in patients with a distal biceps tendon rupture.

Purpose**:** The primary aim of this study is to find a (combination of) test(s) that serves best to diagnose a complete rupture with certainty in the acute phase (≤ 1 month) without missing any complete ruptures. The secondary aims are to determine the best (combination of) test(s) to identify a chronic (>1 month) rupture of the distal biceps tendon and indicate additional imaging in case partial ruptures or tendinitis are suspected.

Study Design: Prospective cohort study.

Methods: 86 patients presenting with anterior elbow complaints or suspected distal biceps injury underwent standardized physical examination, including the Hook test, Passive Forearm Pronation (PFP) test, Biceps Crease Interval (BCI) and Biceps Crease Ratio (BCR). Diagnosis was confirmed intra-operatively (68 cases), by magnetic resonance imaging (MRI) (13 cases) or ultrasound (5 cases).

Results: A combination of the Hook test and BCI (i.e., both tests are positive) is most accurate for both acute and chronic ruptures, however with a different purpose. For acute complete ruptures sensitivity was 94% and specificity 100%. In chronic cases, specificity was also 100%. Weakness on active supination and palpation of the footprint provide excellent sensitivity of 100% for chronic complete ruptures and partial ruptures respectively.

Conclusion**:** The combination of a positive Hook test and BCI serves best to accurately diagnose both acute and complete ruptures of the distal biceps tendon. Weakness on active supination and pain on palpation of the footprint provide excellent sensitivity for chronic complete ruptures and partial ruptures. Using these tests in all suspected distal biceps ruptures allows a physician to omit from imaging for a diagnostic purpose in certain cases, in order to limit treatment delay and hereby provide better treatment outcome, as well as to avoid hospital and social costs.

What is known about the subject: To date, several studies have been published on physical examination tests for complete distal biceps ruptures. These include 5 studies on individual tests (Biceps Squeeze, Hook test, Supination-Pronation test, Bicipital Aponeurosis flex test, BCI and BCR) an 1 on a combination of 3 tests (Hook test, PFP test, BCI), with 5 to 48 patients included. None of the studies on complete ruptures distinguished whether the diagnostic test was performed on acute (≤1 month) or chronic cases. This is essential, since local pathology and therefore the findings at physical examination differ between these two situations. In chronic ruptures, scar tissue may be mistaken for an intact distal biceps and may therefore influence diagnostic accuracy of the physical examination. At the start of our study, no studies were published on partial ruptures or tendinitis.

What this study adds to existing knowledge: This study was performed on the largest prospective sample (86 patients) with suspected distal biceps pathology. We were able to provide diagnostic accuracy to set the correct diagnosis for both acute and chronic ruptures using a combination of physical examination tests with a high diagnostic accuracy. Based on our results we propose a diagnostic algorithm to allow the physician to omit from imaging for a diagnostic purpose in certain cases, in order to limit treatment delay and hereby provide better treatment outcome, as well as to avoid hospital and social costs.

Introduction

Distal biceps tendon ruptures are mostly caused by a sudden extension force on the flexed elbow. Recent literature shows an increasing incidence, while the age of patients with a distal biceps rupture is decreasing[77]. In the acute setting, it is essential to differentiate between a complete and partial rupture, as complete ruptures may require surgical fixation on a short notice (within 2 to 4 weeks[16, 76, 125]). Non-operative management of a complete rupture results in a permanent reduction of flexion- and supination strength[109]. Delayed operative treatment of a complete rupture compromises the ability to perform primary repair, due to retraction of the tendon and peritendinous fibrotic scar tissue. This generally necessitates extended reconstructive surgery, which is related to a higher number of surgical complications[3, 76, 86, 125]. Additional imaging such as magnetic resonance imaging (MRI) or ultrasound is frequently used in case of uncertainty, physical examination is a less invasive, less costly and less time-consuming option. MRI is considered the gold standard in diagnosing distal biceps tendon injuries. However, recent research showed that ultrasound is equally reliable[40]. A quick diagnosis and surgery is a less important matter in partial ruptures and tendinitis, as there is no retraction of the tendon. However, clinical suspicion based on physical examination in partial ruptures is important, since MRI has a low sensitivity (59.1%) but a high specificity (100%)[49]. Accurate physical examination may serve to differentiate between complete and partial ruptures and tendinitis of the distal biceps tendon, leading to optimal treatment decision and indication for additional imaging.

Several retrospective studies have been published on physical examination tests for complete distal biceps ruptures, using MRI and/or surgery as reference test[41, 46, 47, 99, 111, 131].

At the start of our study, no studies were published on partial ruptures or tendinitis. None of the studies on complete ruptures distinguished whether the diagnostic test was performed on acute (≤1 month) or chronic cases. Since local pathology and findings at physical examination may be different, this differentiation is essential as it involves a different treatment regimen. For example, in chronic ruptures scar tissue may be mistaken for an intact distal biceps and influence diagnostic accuracy of the physical examination. In the acute setting, pain and swelling may also mislead the clinician in his or her assessment of the integrity of the distal biceps.
Furthermore, the diagnostic accuracy of physical examination tests used in previous literature is questionable, since all studies are based on a low number of patients and are retrospective in design, potentially causing spectrum bias. This may result in an overestimation of diagnostic accuracy by means of the positive predictive value (PPV) and negative predictive value (NPV)[116]. Therefore, a prospective study design is preferred to avoid bias and also include patients that do not suffer from the disease of interest.

Subsequent to these requirements, a prospective study was designed with the primary purpose to discover a (combination of) test(s) that serves best to diagnose a complete rupture in the acute phase (≤ 1 month) without missing any complete ruptures. Our secondary aims are to determine the best (combination of) test(s) to confirm a chronic (>1 month) rupture of the distal biceps tendon and set the indication for additional imaging in cases that are potentially considering partial ruptures or tendinitis. Ultimately, this data may serve a diagnostic algorithm for clinical use.

Methods

Patient selection

A prospective cohort of consecutive patients presenting with anterior elbow complaints or suspected distal biceps injury were included in our study. Inclusion took place in in the outpatient clinics and emergency departments of two large teaching hospitals between January 2017 and July 2020. Patients were excluded in case of penetrating trauma or fracture, insufficient knowledge of Dutch or English language or significant cognitive impairment. Ethical approval was not required as no additional interventions are performed other than standard care.

Data collection

Physical examination was performed once on every patient by 4 experienced orthopedic upper limb surgeons. The following data was collected using a standard elbow questionnaire: gender, age, hand dominance, affected arm, trauma and duration of complaints (in months). Physical examination consisted of general elbow examination and specific examination of the distal biceps, based on literature[159]. General examination included: carrying angle (normal, valgus, varus), flexion-extension and supination-pronation elbow range of motion.

Index tests

Specific examination consisted of the following items: palpation of the distal biceps footprint on the radial tuberosity, active supination against resistance, the Hook test for integrity of the tendon and for pain[111], the Passive Forearm Pronation (PFP) test for integrity and for pain[41, 66] and the Biceps Crease Interval (BCI) and Biceps Crease Ratio (BCR)[47]. Description of general examination and specific examination tests are provided in table 1.

Table 1 General examination and physical examination tests

General examination	
Carrying angle (normal/valgus/varus)	
Range of motion (normal or decreased compared to contralateral side)	
	Flexion-extension
	Supination-pronation
Specific tests	
Hook test[111]	
Patient position	Seated, passive supination forearm, 90° elbow flexion
Examiner position	Index finger on antecubital fossa
Test	Hook index finger under intact biceps tendon from lateral side

Assessment	No cord-like structure to hook a finger indicates total distal biceps rupture, painful test indicates partial rupture
Passive forearm pronation test[41, 66]	
Patient position	Seated, 90° elbow flexion
Examiner position	Hand on biceps muscle belly, fixate wrist
Test	Palpate biceps muscle belly while pro-/supinating forearm passively
Assessment	No proximal excursion of biceps muscle belly in supination and distal migration in pronation indicates total distal biceps rupture, painful test indicates partial rupture
Biceps crease interval (BCI)[47]	
Patient position	Seated, 90° elbow flexion
Examiner position	Fixate wrist, index finger on antecubital fossa
Test	Passively extend the elbow, supinate forearm. Mark flexion crease in antecubital fossa. Mark start of the biceps curve. Measure distance between marks
Assessment	Absolute BCI value >6cm indicates complete distal biceps rupture
Biceps crease ratio (BCR)[47]	
Patient position	Seated, 90° elbow flexion
Examiner position	Fixate wrist, index finger on antecubital fossa
Test	Repeat steps of BCI test on contralateral arm, calculate ratio between BCI's in both arms
Assessment	BCR >1.2 indicates complete distal biceps rupture

Reference test

Based on recommendations in literature, either MRI (Ingenia–3T; Phillips Healthcare), ultrasound (GE Logiq E9, GE Healthcare with a multi-frequency linear transducer) and/or confirmative surgery was performed on each patient[49, 93, 134]. Surgery was only performed in cases where the patient and surgeon consented on this treatment and not for diagnostic purposes only. Intra-operative findings are considered most accurate and overrule a diagnosis established by imaging. Additional imaging was performed and/or assessed by experienced musculoskeletal radiologists. The examining orthopedic surgeon was blinded for the radiology report and images in case diagnostic imaging was performed before patients' visit to the outpatient clinic. Intra-operative assessment was performed by the aforementioned 4 experienced orthopedic surgeons, who were not blinded for physical examination and/or additional imaging results. A complete rupture was defined as complete loss of distal biceps tendon fibers attached to the radial tuberosity. In case a significant part of the tendon fibers remained intact that prevented retraction, the rupture was classified as partial. Cases with a complete rupture of the distal biceps tendon but an intact lacertus fibrosis were also classified as complete distal biceps tendon ruptures. Tendinitis was defined as inflammation around the tendon or bicipitoradial bursa.

Statistical analysis

Data-analysis was performed using Statistical Package for the Social Sciences 26 (IBM Corporation, Armonk, NY, USA). Physical examination tests were analyzed separately and in combination to determine the diagnostic accuracy for the 3 targeted diseases: complete distal biceps tendon rupture, partial distal biceps tendon rupture and tendinitis of the distal biceps tendon. Both acute (≤1month), chronic (>1 month) and total group were assessed. Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) and likelihood ratios (LR) were calculated using 2x2 tables with accompanying 95% confidence intervals (CI). A positive likelihood ratio (LR+, i.e. post-test probability of disease presence) >10 and a negative likelihood ratio (LR−, i.e. post-test probability of disease absence) <0.1 were defined as large[62].

Results

Eighty-six consecutive patients met the inclusion criteria and had a complete work-up. Nearly half of the patients (n = 42, 49% of the total group) had a complete distal biceps rupture, 29 patients (34%) a partial rupture, 10 (12%) tendinitis and 5 (6%) another diagnosis.

Demographics

Majority of the patients were male (n = 78, 91%). Age varied from 24 to 74 y/o, with a mean of 49.3 y/o (95% CI 47.3-51.3). Hand dominance was in line with the general population, 86%

of the patients were right-dominant. In 65% of the cases the dominant side was affected. The median duration of complaints in the total sample was 6.0 months (range 0-84 months), with 23 patients (27%) presenting within the first month. Amongst patients with a complete rupture, 21 (50%) presented within the first month, the other half had chronic complaints. Additional imaging was performed in 68 patients (79%). Of these, 42 had an MRI, 20 an ultrasound and in 6 patients both MRI and ultrasound were performed. The remaining 18 patients (21%) underwent surgery without additional diagnostic imaging. In 68 patients (79%) the final diagnosis was confirmed by surgery, 13 (15%) diagnoses were based on MRI findings and in 5 patients (6%) diagnosis was only confirmed using ultrasound, all suspected and confirmed either partial rupture or tendinitis.

Physical examination tests

Pain on palpation of the footprint on the radial tuberosity and weakness with active supination (in 90 degrees of flexion) are sensitive (91% and 95%, respectively) for complete distal biceps ruptures, weakness with active supination being 100% sensitive in chronic complete ruptures. However, specificity (11% and 21%, respectively), predictive values and likelihood ratios are low.

The Hook test has moderate sensitivity (71%), however in acute injuries (<1 month) sensitivity is higher (86%). This test is able to correctly reject patients without a complete rupture, with a specificity of 95% overall and 100% in acute cases.

The BCI and BCR showed the same result in all cases. Sensitivity is 81% overall, rising to 86% in acute cases. Specificity was higher in chronic cases (93%) compared to acute injury (50%).

The PFP test demonstrates weaker diagnostic accuracy compared to the Hook test and BCI/BCR. It is less accurate in the acute setting compared to chronic ruptures with a sensitivity of 74% overall, 67% in acute cases and 81% in chronic cases. Specificity is 77%, 50% and 79% for overall, acute and chronic cases respectively.

Combining the Hook test and BCI (i.e., both tests are positive vs. none positive) sensitivity is 94% in acute cases, with a specificity and PPV of 100%. For chronic ruptures this combination is less sensitive (71%), however a specificity and PPV of 100% was observed. Adding the PFP test provides a higher sensitivity in the chronic setting: if 2 out of 3 tests are positive (vs. none positive), sensitivity is 80%, specificity 97% and PPV 94%. However, in order to select a patient with a chronic distal biceps tendon rupture for reconstruction surgery, a specificity of 100% is desired and therefore the combination of the Hook test and BCI is most useful.

Palpation of the footprint on the radial tuberosity and weakness with active supination both have high sensitivity (100% and 93%, respectively) for partial distal biceps rupture. However, specificity (16% and 10%, respectively), predictive values and likelihood ratios are low.
A painful Hook test and PFP test show moderate sensitivity for identifying a partial rupture (76 and 72%, respectively), with low specificity, PPV/NPV and likelihood ratios. Combining these tests leads to higher sensitivity (up to 85%), however none reached the sensitivity of solely palpation of the footprint. In patients with distal biceps tendinitis, the values were low for all tests.

Diagnostic accuracy of individual tests and combinations are summarized in table 2.
Table 2 – Diagnostic accuracy of physical examination tests (individually and combined)

Condition	Test	Sens. (%) (95% CI)	Spec. (%) (95% CI)	PPV (%) (95% CI)	NPV (%) (95% CI)	LR+	LR-
Complete rupture (total) (n=42)	Palpation footprint	90.5 (76.5-96.9)	11.4 (4.3-25.4)	49.4 (37.9-60.9)	55.6 (22.7-84.7)	1.02	0.84
	Active supination	95.2 (82.6-99.2)	20.5 (10.3-35.8)	53.3 (41.5-64.8)	81.8 (47.8-96.8)	1.20	0.23
	Hook test	71.4 (55.2-83.8)	95.5 (83.3-99.2)	93.8 (77.8-98.9)	77.8 (64.1-87.5)	15.71	0.30
	PFP test	73.8 (57.7-85.6)	77.3 (61.8-88.0)	75.6 (59.4-87.1)	75.6 (60.1-86.6)	3.25	0.34
	BCI	81.0 (65.4-90.9)	90.9 (77.4-97.0)	89.5 (74.3-96.6)	83.3 (69.2-92.0)	8.9	0.21
	Combi Hook + BCI (both positive)	82.4 (64.8-92.6)	100 (88.6-100)	100 (85.0-100)	86.4 (72.0-94.3)	Infinity	0.18
	Combi Hook + BCI + PFP (2 out of 3 positive)	87.2 (71.8-95.2)	93.8 (77.8-98.9)	94.4 (80.0-99.0)	85.7 (69.0-94.6)	13.95	0.14
Complete rupture (acute, <1m) (n=21)	Palpation footprint	90.5 (76.5-96.9)	11.4 (4.3-25.4)	49.4 (37.9-60.9)	55.6 (22.7-84.7)	1.02	0.84
	Active supination	90.5 (68.2-98.3)	50.0 (2.7-97.3)	95.0 (73.1-99.7)	33.3 (1.8-87.5)	1.81	0.19
	Hook test	85.7 (62.6-96.2)	100 (19.8-100)	100 (78.1-100)	40.0 (7.3-83.0)	Infinity	0.14
	PFP test	66.7 (43.1-84.5)	50.0 (2.7-97.3)	93.3 (66.0-99.7)	12.5 (0.7-53.3)	1.33	0.66
	BCI	85.7 (62.6-96.2)	50.0 (2.7-97.3)	94.7 (71.9-99.7)	25.0 (1.3-78.1)	1.71	0.29
	Combi Hook + BCI (both positive)	94.1 (69.2-99.7)	100 (5.4-100)	100 (75.9-100)	50.0 (2.7-97.3)	Infinity	0.06
	Combi Hook + BCI + PFP (2 out of 3 positive)	94.7 (71.9-99.7)	50.0 (2.7-97.3)	94.7 (71.999.7)	50.0 (2.7-97.3)	1.89	0.11
Complete rupture (chronic, >1m) (n=21)	Palpation footprint	90.5 (76.5-96.9)	11.4 (4.3-25.4)	49.4 (37.9-60.9)	55.6 (22.7-84.7)	1.02	0.84
	Active supination	100 (80.8-100)	19.0 (9.1-34.6)	38.2 (25.7-52.3)	100 (59.8-100)	1.24	0
	Hook test	57.1 (34.4-77.4)	95.2 (82.6-99.2)	85.7 (56.2-97.5)	81.6 (67.5-90.8)	12	0.45
	PFP test	81.0 (57.4-93.7)	78.6 (62.8-89.2)	65.4 (44.4-82.1)	89.2 (73.6-96.5)	3.78	0.24
	BCI	76.2 (52.5-90.9)	92.9 (79.4-98.1)	84.2 (59.5-95.8)	88.6 (74.6-95.7)	10.67	0.26
	Combi Hook + BCI (both positive)	70.6 (44.0-88.6)	100 (88.3-100)	100 (69.9-100)	88.1 (73.6-95.5)	Infinity	0.29
	Combi Hook + BCI + PFP (2 out of 3 positive)	80.0 (55.7-93.3)	96.7 (80.9-99.8)	94.1 (69.2-99.7)	87.9 (70.9-96.0)	24	0.21
Partial rupture (n=29)	Palpation footprint	100 (85.4-100)	15.8 (7.9-28.4)	37.7 (27.1-49.5)	100 (62.9-100)	1.19	0
	Active supination	93.1 (75.8-98.8)	10.3 (5.1-19.2)	25.7 (17.9-35.3)	81.8 (47.8-96.8)	1.04	0.67
	Hook test for pain	75.9 (56.1-89.0)	49.1 (35.8-62.6)	43.1 (29.6-57.7)	80.0 (62.5-90.9)	1.49	0.49
	PFP test for pain	72.4 (52.5-86.6)	47.4 (34.2-60.9)	41.1 (27.9-55.8)	77.1 (59.4-89.0)	1.38	0.58
	Combi Hook pain + PFP pain (both positive)	85.0 (61.1-96.0)	47.5 (31.8-63.7)	44.7 (29.0-61.5)	86.4 (64.0-96.4)	1.62	0.32
Tendinitis (n=10)	Palpation footprint	70.0 (35.4-91.9)	7.9 (3.3-17.0)	9.1 (4.0-18.4)	66.7 (30.9-91.0)	0.76	3.8
	Active supination	60.0 (27.4-86.3)	9.2 (4.1-18.6)	8.0 (3.3-17.2)	63.6 (31.6-87.6)	0.66	4.34
	Hook test for pain	50.0 (20.1-79.9)	39.5 (28.7-51.4)	9.8 (3.7-22.2)	85.7 (69.0-94.6)	0.83	1.27
	PFP test for pain	40.0 (13.7-72.6)	38.2 (27.5-50.1)	7.8 (2.5-19.7)	82.9 (65.7-92.8)	0.65	1.57

Discussion

This study on the largest prospective series up to date in 86 patients demonstrates that a combination of a positive Hook test and BCI is most accurate for both acute and chronic ruptures. For acute ruptures a test with well-balanced sensitivity and specificity is essential, in order to rule out anything but a complete rupture with certainty (specificity 100%) without missing any (sensitivity 94%). For chronic ruptures, it is important to rule out anything but a complete rupture with certainty in order to decide on operative treatment (specificity 100%). Weakness on active supination and palpation of the footprint provide excellent sensitivity (100%) for chronic complete ruptures and partial ruptures respectively. For tendinitis none of the tests provided sufficient diagnostic accuracy to rule in or rule out disease. Using Hook test and BCI combined in all suspected distal biceps ruptures allows physicians to omit from imaging for the purpose of diagnosing in certain cases, in order to limit treatment delay and hereby provide better treatment outcome, as well as avoid hospital and social costs. Based on our findings, we propose a diagnostic algorithm in figure 1.

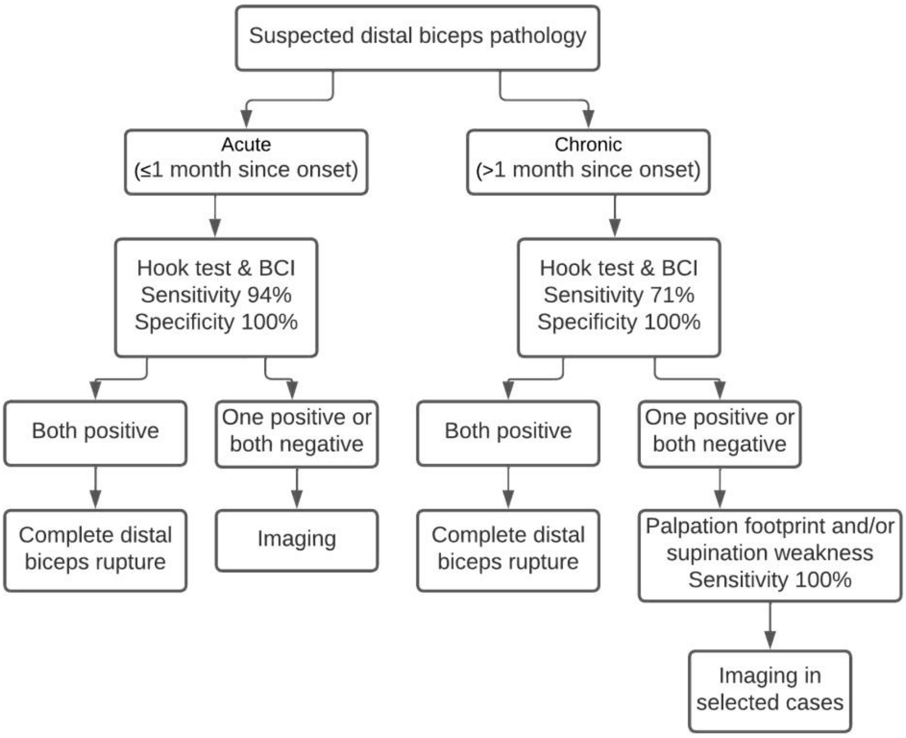


Figure 1: Proposed diagnostic algorithm for suspected distal biceps injuries

In previous work we assessed the methodologic quality of all studies on distal biceps pathology published until 2016, identifying bias and applicability concerns[159]. These concerns are mainly caused by patient selection: half of the studies were case-control design[47, 99, 111] and all studies had a high prevalence of disease (73 to 100%) with a low number of patients (5 to 48). Recently a study investigated the sensitivity of the Hook test in a larger sample (202 cases) with a lower prevalence of complete ruptures. However a retrospective design was used, potentially introducing selection bias[92]. Also, most studies focused on individual tests for (acute) complete distal biceps tendon ruptures rather than a combination of tests or other entities such as partial ruptures, chronic ruptures or tendinitis.

The Hook test, PFP test, BCI/BCR, Biceps Squeeze test and Supination-pronation test were investigated for individual use and showed a sensitivity ranging from 81-100% and specificity from 50-100%[47, 92, 99, 111, 131]. One study tested a combination of tests: the Hook test, PFP and BCI combined reached a sensitivity of 100%, specificity of 50%[41]. In comparison to our results, previous literature on individual tests for complete distal biceps ruptures showed similar diagnostic accuracy. For the combination of tests, the sensitivity in the combinations we investigated was slightly lower: 94% and 80% for acute and chronic ruptures respectively in our study, compared to 100% overall in the study by Devereaux et al[41]. For specificity however, our sample reached 100% and 97% for acute and chronic ruptures respectively, compared to 50% by Deveraux et al. However, a direct comparison cannot be made since none of the previous studies differentiated diagnostic accuracy of tests between acute and chronic ruptures.

In all studies, either MRI and/or intra-operative findings were used as reference test. MRI is the gold standard to diagnose a complete distal biceps tendon rupture, however recent research showed that ultrasound is equally reliable[40]. The sensitivity and specificity of an MRI for complete tears are 100% and 82.8%, respectively[49]. Recent literature reported no significant differences in sensitivity and specificity in detecting partial distal biceps injuries when FABS view MRI and standard MRI were compared[134]. Ultrasound was found equally accurate in cases of complete and high-grade partial distal biceps tendon ruptures[40]. For partial tears, the overall accuracy rate of MRI (66.7%) was the same as ultrasound, however MRI has a low sensitivity (59.1%) but a high specificity (100%)[49, 93].

The main strengths of our study are provided by the fact that it is the largest prospectively collected sample of patients with distal biceps pathology, consisting of both patients with acute and chronic complete and partial ruptures, as well as tendinitis and other anterior elbow complaints. Using a combination of tests provided even more reliable results, with the ability to more accurately discern between acute and chronic complete ruptures. Therefore, we believe that our results are applicable and valuable for clinical practice.

The distal biceps tendon rupture is an increasingly popular topic for research. During the inclusion period of our study two novel tests for partial ruptures and tendinitis were published. Shim et al presented a novel clinical test for partial tears of the distal biceps tendon, the TILT sign[141]. This test however does not add to our study since it is performed exactly the same way as we performed palpation of the radial tuberosity. Caekebeke et al introduced the biceps provocation test (BPT) for partial ruptures and tendinitis, with a sensitivity and specificity of both 100%. This test includes 2 parts: first resisted flexion of the elbow with the forearm supinated and secondly with the forearm pronated[28]. The generalizability of this test is questionable since the authors performed the test only on patients with a partial rupture (and healthy participants) and not on patients with a complete rupture. In order to use this test to differentiate between complete and partial ruptures, further research is needed.

Furthermore, we did not include the Supination Pronation test, because the only difference between this test and the PFP test is the active motion of the patient (without testing the strength) versus passive motion[99]. Both tests were designed to test the integrity of the tendon and accompanying muscle belly, so in our experience active motion has limited added value. However, we incorporated testing the supination strength against resistance to test active motion and strength.

Another limitation is the fact that we did not perform inter- or intra-rater reliability. ElMaraghy et al. investigated the inter-rater reliability for the BCI test observing an intraclass correlation coefficient (ICC) of 0.79[47]. For all other tests no inter- or intra-rater reliability was available in literature, leaving this subject for further research.

Conclusion

The combination of a positive Hook test and BCI is highly sensitive (94%) and specific (100%) to diagnose acute, complete ruptures of the distal biceps tendon. A similar high specificity (100%) was found for identifying chronic rupture of the distal biceps, when the combination of these tests was performed. Weakness on active supination and pain on palpation of the footprint provide excellent sensitivity (100%) for chronic complete ruptures and partial ruptures, however physical examination tests are not accurate enough to confirm either partial rupture nor tendinitis, leaving an indication for additional diagnostic imaging in these cases.

Chapter 6

Clinical diagnosis of lateral sided elbow pain: predictors to recognize a diagnosis other than a tennis elbow

Elisa L. Zwerus MD[1,2], Renée Keijsers MD[3], Joost W. Colaris MD PhD[1], Bertram The MD PhD[4], Michel P.J. van den Bekerom MD PhD[5,6], Denise Eygendaal MD PhD[1]

[1]Department of Orthopaedic Surgery & Sports Medicine, Erasmus University Medical Centre, Rotterdam, The Netherlands
[2]Department of Orthopaedic Surgery, Zuyderland MC, Sittard-Heerlen, The Netherlands
[3]Department of Orthopaedic Surgery, Sint Maartenskliniek, Nijmegen, The Netherlands
[4]Department of Orthopaedic Surgery, Amphia, Breda, The Netherlands
[5]Shoulder and elbow unit, Department of Orthopaedic Surgery, OLVG, Amsterdam, The Netherlands
[6]Department of Human Movement Sciences, Vrije Universiteit, Amsterdam, The Netherlands

Manuscript in preparation for submission.

Abstract

Background. Lateral-sided elbow pain is most frequently diagnosed as lateral epicondylitis. It is diagnosed based on symptoms and physical examination only, eliminating the need for diagnostic imaging in the initial workup. However, up to 25% of patients with lateral elbow pain are misdiagnosed as lateral epicondylitis. Various conditions produce similar symptoms to lateral epicondylitis and should be recognized to provide a correct and timely diagnosis and treatment. Our study aims to identify items in patient characteristics, history and physical examination that predict a diagnosis other than lateral epicondylitis in adults with lateral elbow pain.

Methods. A prospective cohort of 170 patients presenting to our outpatient orthopedic clinics with lateral-sided elbow complaints was included. All patients underwent a

standardized diagnostic protocol. Bivariable analysis and multivariable binary logistic regression with a stepwise backward selection procedure were used to identify variables associated with a diagnosis other than lateral epicondylitis.
Results. Almost half of the patients (46.5%) were diagnosed with a diagnosis other than lateral epicondylitis. Age ≤30 ($p<0.001$), acute onset ($p = 0.045$), locking ($p<0.001$), hydrops ($p<0.001$), any positive instability test ($p = 0.013$) and a negative Maudsley test ($p<0.001$) were found independent factors to predict a diagnosis other than lateral epicondylitis.
Conclusion. Independent predictors were identified for use in the clinical diagnosis of lateral-sided elbow pain, aiming to help physicians recognize a diagnosis other than lateral epicondylitis.

Introduction

Lateral-sided elbow pain is most frequently diagnosed as lateral epicondylitis, also known as tennis elbow. Lateral epicondylitis is diagnosed based on symptoms and physical examination only, eliminating the need for diagnostic imaging in the initial workup. However, everal other causes of lateral elbow pain should be recognized, especially when symptoms persist[82]. Up to 25% of patients with lateral elbow pain are misdiagnosed as lateral epicondylitis[130]. A variety of conditions produce similar symptoms to lateral epicondylitis, including radial tunnel syndrome (also known as posterior interosseus nerve syndrome or supinator syndrome), synovial fringe (radio-capitellar or posterolateral synovial plica), posterolateral rotatory instability (PLRI), osteochondritis dissecans (OCD), Panner's disease, degenerative or inflammatory arthropathy, occult fractures or pathology in surrounding joints such as cervical radiculopathy. Intra-articular pathology has been reported in up to 86% during arthroscopic surgery for refractory lateral epicondylitis cases[7]. Lateral ligament instability has also been reported as a coexisting factor in 46-64% of cases with refractory lateral epicondylitis[8]. A conventional radiograph is usually used to exclude other pathology at an outpatient visit. However these do not seem to be of added value in patients without joint crepitus, instability, deformity, loss of motion or a prior history of elbow fracture, dislocation, surgery, or arthritis[120].

To date, no studies report on the diagnostic accuracy of history items for diagnoses causing lateral-sided elbow pain. A few studies were published on physical examination tests, including one study on lateral epicondylitis using the grip strength dynamometer[43] and two small studies (eight patients each) on four different tests for PLRI[9, 126]. No studies to date describe the diagnostic accuracy of physical examination tests for radial tunnel syndrome, synovial fringe, OCD, Panner's disease, or degenerative or inflammatory arthropathy.

A systematic approach to diagnosing a patient with lateral elbow pain is desired to recognize causes other than lateral epicondylitis. Therefore, our study aims to identify patient characteristics, history and physical examination items that predict a diagnosis other than lateral epicondylitis in adults with lateral elbow pain.

Methods

Patient selection

A prospective cohort of patients presenting to our outpatient orthopedic clinics with lateral sided elbow complaints was included in our study. Inclusion occurred in three large teaching hospitals between January 2017 and July 2022. Follow-up to identify the definitive diagnosis was carried out until April 2024. Patients were excluded in case of penetrating trauma or fracture, insufficient knowledge of Dutch or English language or significant cognitive impairment. The local ethical committee waived ethical approval; patients were not required to provide informed consent because no additional actions were performed besides standard care.

Data collection

Patients underwent a standardized diagnostic protocol performed by three experienced upper extremity orthopedic surgeons (BT, DE, MB). Patient characteristics, history items and physical examination test are listed in table 1. Execution of these tests is either easily comprehensible or based on detailed descriptions in previous literature[159].

TABLE 1: Protocol including patient characteristics, history items and physical examination tests

Patient characteristics and history items	Physical examination tests
Gender Age (years) Duration of complaints (months) Dominant hand affected Manual labor Hand/arm related sports Onset: acute/gradual Locking Crepitations Neurologic symptoms	Alignment Motion restriction Joint effusion Grip & grind test passive/active in 90° Any instability test (Varus stress test, Stand-up test, Table top relocation test, Drawer test, Pivot shift test) Palpitation capitellum (pain) Palpation lateral epicondyle (pain) Mills test Maudsley test Cozen test

Reference test

Patients were diagnosed with lateral epicondylitis based on additional imaging. When not available, according to the protocol proposed by Dorf et al[43]. Patients were required to have tenderness over the ECRB or the common extensor origin and at least 2 of the three following criteria:

(1) pain with resisted wrist extension (Cozen's test),

(2) pain with resisted middle finger extension (Maudsley's test),

(3) pain with the elbow extended and the wrist flexed and pronated (Mills test).

In addition, every patient had a conventional X-ray (AP and lateral) of the elbow. The orthopedic surgeon initiated the discission to perform additional imaging and/or a surgical procedure, based on the probable diagnosis and/or differential diagnosis. For soft-tissue conditions, ultrasound or magnetic resonance imaging (MRI) with or without arthrography was performed, and bony conditions were confirmed using computed tomography (CT). In cases where surgery was performed, per-operative findings overruled a diagnosis based on imaging. The intra-operative assessment was performed by the aforementioned three experienced upper extremity orthopedic surgeons, who were not blinded for physical examination and/or additional imaging results.

Statistical analysis

Data-analysis was performed using Microsoft Excel 2010 (Microsoft Corp. Washington, USA) and Statistical Package for the Social Sciences 29 (IBM Corporation, Armonk, NY, USA). Descriptive statistics were calculated: median and range for non-normally distributed continuous variables, means and standard deviations for normally distributed continuous variables, and frequencies and percentages for categorical variables.

Bivariable analysis was performed to assess if any variables within patient characteristics, history items and physical examination were associated with the definitive diagnosis other than lateral epicondylitis. A χ^2 test was used for categorical variables, and items with five or fewer items were removed.

Subsequently, variables with a P-value <0.1 entered a multivariable binary logistic regression with a stepwise backward selection procedure. At each step, the variable with the largest P-value was eliminated. This process was repeated until all variables in the equation reached a P-value <0.05. Multivariable binary logistic regression was limited to 10 events per variable. Sensitivity, specificity, Positive Predictive Value (PPV) and Negative Predictive Value (NPV) were calculated for tests targeting the diagnosis of lateral epicondylitis.

Results

One hundred seventy patients with lateral elbow pain were included. Ninety-one patients were male (53.5%). Age ranged from 9 to 79 with a median of 45 y/o. The median of the duration of complaints was 12 months (range 1-240). The dominant side was affected in 62.4% (n = 106). Hand-related sports were performed in 41.8% (n = 71) of patients, and hand labour or desk work in 75.9% (n = 129). Most patients received treatment before they visited the orthopedic clinic: 101 patients (59.4%) were treated by a physiotherapist, and 35 patients (20.5%) received one or more corticosteroid injections.

Ninety-seven patients (57.1%) underwent additional imaging and/or surgical treatment; the remaining 73 patients were diagnosed using the physical examination protocol described in the methods section. The final diagnoses are listed in Table 2.

TABLE 2: Final diagnosis

Diagnosis	N	% of total
Lateral epicondylitis	91	53.5
Other diagnosis	79	46.5
OCD	36	21.2
LCL injury	18	10.6
Arthrosis	9	5.3
Supinator syndrome	7	4.1
Synovial fringe	4	2.4
Other	5	2.9
Total	137	100

Bivariable analysis of patient characteristics, history- and physical examination items demonstrated that age ≤30, duration ≤12 months, hand/arm related work, hand/arm related sports, acute onset, locking, crepitus, malalignment, hydrops, F/E motion restriction, positive passive G&G test, positive results on any of the instability tests, pain on palpitation capitellum, absence of pain on palpitation epicondyle, negative Mills, Maudsley, and Cozen test were associated with a diagnosis other than lateral epicondylitis (table 3). Multivariable analyses subsequently identified age ≤30, acute onset, locking, hydrops, positive results on any instability test, and a negative Maudsley test as independent predictors (table 4).

TABLE 3: Bivariable analysis of patient characteristics, history items and physical examination.

Variable	LE	Other	P-value
Patient characteristics and history items			
Sex			0.146
Male	47	32	
Female	44	47	
Age			<0.001*

≤30 >30	4 87	50 29	
Duration ≤12 months >12 months	 66 25	 38 41	<0.001*
Affected side Dominant Non-dominant	 57 34	 49 30	0.935
Work Hand/arm related Not hand/arm related	 64 27	 65 14	0.069
Sport/hobby Hand/arm related Not hand/arm related	 31 39	 40 60	0.029*
Onset Acute Gradual	 18 73	 32 47	0.003*
Locking Yes No	 0 91	 48 31	<0.001*
Crepitus Yes No	 2 89	 16 63	<0.001*
Neurology Yes No	 16 75	 22 57	0.109
Physical examination			
Alignment Varus or valgus malalignment Normal alignment	 1 90	 16 63	<0.001*
Hydrops			<0.001*

Yes No	0 91	32 47	
F/E motion Restricted Normal (compared to contra-lateral)	 5 86	 45 34	<0.001*
G&G passive Positive Negative	 4 70	 22 35	<0.001*
G&G active Positive Negative	 41 33	 28 29	0.475
Any instability test Increased laxity Stable	 0 89	 21 57	<0.001*
Pain on palpitation capitellum Positive Negative	 12 77	 44 34	<0.001*
Pain with palpitation lateral epicondyle Positive Negative	 84 7	 20 59	<0.001*
Mill's test Positive Negative	 64 27	 12 67	<0.001*
Maudsley test Positive Negative	 82 9	 12 67	<0.001*
Cozen test Positive Negative	 78 13	 15 64	<0.001*

** $\chi 2$ test was significant at $P < 0.05$*

TABLE 4: Multivariable Logistic Regression Analysis of patient characteristics, history- and physical examination items.

Variable	Odds ratio for other diagnosis (95% CI)	P-value
Age ≤30	37.5 (12.5-112.9)	<0.001
Acute onset	2.8 (1.4-5.5)	0.045
Locking	3.9 (2.9-5.3)	<0.001
Hydrops	2.9 (2.3-3.7)	<0.001
Instability tests	2.6 (2.1-3.1)	0.013
Negative Maudsley test	50.9 (20.2-128.0)	<0.001

Discussion

This study presented factors in patient characteristics, history and physical examination that predict a diagnosis other than lateral epicondylitis in adults with lateral elbow pain. Our prospective cohort of 170 patients all underwent a standardized diagnostic protocol, 46.5% of the patients were diagnosed with a diagnosis other than lateral epicondylitis. Age ≤30 ($p<0.001$), acute onset ($p = 0.045$), locking ($p<0.001$), hydrops ($p<0.001$), any positive instability test ($p = 0.013$) and a negative Maudsley test ($p<0.001$) were found independent predictors for a diagnosis other than lateral epicondylitis. These findings can help physicians to recognize a diagnosis other than lateral epicondylitis.

In previous literature, no studies were found to identify patient characteristics, history, or physical examinations that predicted a diagnosis other than lateral epicondylitis. Furthermore, we did not find any studies on the diagnostic accuracy of patient characteristics or history items. A few studies were published on physical examination tests for lateral-sided elbow pain. The fact that the diagnosis of lateral epicondylitis is based on clinical findings makes it challenging to draw conclusions on the diagnostic accuracy of physical examination tests, since there is no explicit reference standard. The only test for lateral epicondylitis investigated was the grip strength test[43]. For this test, the patient is asked to squeeze a dynamometer as firmly as possible in 90° elbow flexion and full extension. Assessment is based on a 5%, 8% or 10% decrease in grip strength between flexion and extension. Sensitivity is 83-80-78% with a 5-8-10% decrease in strength respectively, and the specificity is 80-85-90%. The reference test used in this study is based on clinical symptoms, including tenderness over the ECRB, Mills' test, Maudsley's test and Cozen's test. Furthermore, two small studies (eight patients each) described tests for PLRI. The table-top relocation test[9], stand-up/chair push-up test[126] and push-up test[126] show a sensitivity from 88 to 100%; specificity was not available since no patients were included that did not suffer from PLRI. The pivot shift test is performed on a fully supinated and extended elbow, followed by valgus stress and axial compression while flexing the elbow. It is positive when the radial head dislocates around 40° flexion. The test shows sensitivity of only 38% in the awake patient but 100% sensitivity when performed under anesthesia[126]. There are no studies to date describing the diagnostic accuracy of physical examination tests for radial tunnel syndrome, synovial fringe, OCD, Panner's disease, or degenerative or inflammatory arthropathy.

Comparing our results with previous studies is not eligible since the study aims and outcome measures differ. The main strength of our study is the design: we have chosen not only to focus on the diagnostic value of individual physical examination tests of one disease but to combine these with demographic and history items to use all the information the patient provides you with. Furthermore, we generated a relatively large sample of patients presenting with different causes of lateral elbow pain that may mimic lateral epicondylitis. Therefore, we believe that our findings are applicable and valuable for clinical practice.

There are some limitations in our study. Firstly, it was hard to define a reference standard since lateral epicondylitis is a clinical diagnosis, and currently, no better alternative is proposed in the literature. However, it is widely adopted in many high-quality randomized clinical trials and has been adopted in studies of outcome measures. To generate the most reliable results possible, we used a formerly described protocol to diagnose lateral epicondylitis[43], and when available additional imaging (initiated by the treating orthopedic surgeon based on the probable diagnosis and/or differential diagnosis) or per-operative findings.

Secondly, the physical examination in our study was performed by experienced upper extremity orthopedic surgeons and may be less accurate when performed by a less experienced person (e.g., a resident or general practitioner). However, the predicting factors mostly consist of demographics and history items and Maudsley's test, which is easy to perform.

The results presented in our study are the first step in developing a model based on factors in patient characteristics, history and physical examination that predict a diagnosis other than lateral epicondylitis in adults with lateral elbow pain. Further research should focus on validating the predictors.

Conclusion

Age ≤30, acute onset, locking, hydrops, any positive instability test and a negative Maudsley test were found to be independent predictors for use in the clinical diagnosis of lateral-sided elbow pain, aiming to help physicians recognize a diagnosis other than lateral epicondylitis.

Chapter 7

Accuracy and quality of educational videos for elbow physical examination: A search from the earliest year until October 2018

Elisa Louise Zwerus[1,3,4], Rik Molenaars[1,2], Michel van den Bekerom[3], Bertram The[4]
Denise Eygendaal[1,4]

[1] Department of Orthopaedic Surgery, Amsterdam UMC, location AMC, Amsterdam, The Netherlands

[2] Harvard Medical School at Massachusetts General Hospital, Sports Medicine Center, Boston, Massachusetts, USA

[3] Shoulder and elbow unit, Department of Orthopaedic Surgery, OLVG, Amsterdam, The Netherlands

[4] Department of Orthopaedic Surgery, Amphia Hospital, Breda, The Netherlands

Arch Bone Jt Surg. 2021 Jan;9(1):93-101. doi: 10.22038/abjs.2020.47221.2310.

Abstract

Introduction: Medical students and residents rely increasingly on web-based education. Online videos provide unique opportunities to share knowledge. The objective of this study was to investigate the accuracy and quality of instructional videos on the physical examination of the elbow and identify factors influencing educational usefulness.

Methods**:** A search on YouTube, VuMedi, Orthobullets, and G9MD was performed. Videos were rated for accuracy and quality by two independent authors using a modified version of a validated scoring system for the nervous and cardiopulmonary system. Inter-rater reliability was analysed.

Results: Twenty-three out of 126 videos were indicated as useful for educational purposes. Accuracy, quality and overall scores were significantly higher for videos from specialized platforms (VuMedi, Orthobullets, G6MD) compared to YouTube. Video accuracy and quality

varied widely and were not correlated. Number of days online, views, and likes showed no or weak correlation with accuracy and quality. For the overall score, our assessment tool showed excellent inter-rater reliability.

Conclusion**:** There is considerable variation in accuracy and quality of currently available online videos on the physical examination of the elbow. We identified 23 educationally useful videos and provided an assessment method for the quality of educational videos. In educational settings, this method may help students to assess video reliability and aid educators in the development of high-quality instructional online content.

Introduction

Driven by increasing emphasis on problem-based and self-directed learning, medical students and doctors in orthopaedic specialty training rely increasingly on the internet (Google, YouTube) as learning resource [23, 75, 80]. As students' or residents' performance on physical examination may be less supervised in comparison to other clinical skills, such as surgical competence, online videos may provide a valuable source for education of physical examination skills. Cognitive psychological research has shown that videos can help viewers to understand techniques and manage the sequential steps of physical examination and approach of patients [6, 30, 51, 96, 145, 149]. YouTube is the largest open-access video sharing platform available with over four billion videos watched every day. YouTube offers a wide variety of user-generated and corporate media videos, including video clips, TV show clips, music videos, short and documentary films, audio recordings, movie trailers, live streams, and other content such as video blogging, short original videos, and educational videos. In this latter category, YouTube provides access to educational videos on a wide variety of orthopaedics-related topics. Compared to YouTube, VuMedi, G9MD, and Orthobullets are online platforms with video content that is more directly focused on orthopaedic topics, requiring user-registration to obtain access.

The accuracy and quality of educational videos for health care providers have been studied for physical examination of the shoulder[83, 151], the nervous system[13], cardiovascular and respiratory systems[14] and direct ophthalmoscopy[21]. Overall, quality and accuracy showed a wide variety among videos, not related to the amount of views or likes. The same conclusion was reached by studies focusing on educational videos for medical and nursing students about subjects such as anatomy, electrocardiography, and pharmacokinetics [2, 11, 12, 26, 29, 123, 139]. This variability in quality and accuracy makes it challenging for students and residents to identify educationally valuable videos for self-learning. A guideline on instruction videos for laparoscopic cholecystectomy was published in 2018, including a list with 45 statements[31]. Unfortunately, most of the statements in this list are not applicable for videos on physical examination skills. Currently, there are no studies on the accuracy and quality of educational videos on physical examination of the elbow.

In this study, the accuracy and quality of videos on general and specific physical examination of the elbow available through YouTube, VuMedi, Orthobullets, and G9MD were assessed using a standardized scoring system. We hypothesized considerable variability in quality and accuracy of the physical examination of the elbow in currently available online videos, with higher quality and accuracy of videos available through specialized platforms (VuMedi, Orthobullets, and G9MD) compared to YouTube, and high inter-rater reliability of quality and accuracy assessment of online videos using the modified scoring system.

Materials and methods

Search

A YouTube search was performed on October 7, 2018 using key words aiming at general physical examination, namely “elbow exam” and “OSCE” (objective structured clinical examination), and key words aiming at specific tests **[Table 1].** YouTube (https://www.youtube.com/) search settings were standard (sorted by relevance) and filtered for individual videos. In addition, all available videos in VuMedi (https://www.vumedi.com/), Orthobullets (https://www.orthobullets.com/) and G9MD (https://g9md.tv/) in the elbow and upper extremity sections were reviewed.

Table 1 – Search terms and hits

Search term	Hits
“Elbow exam”	101000
“Elbow OSCE”	45920
“Hook test biceps”	1690
“Biceps squeeze test”	2050

“Biceps crease interval test”	87
“Biceps crease ratio test”	7
“Triceps squeeze test rupture”	3560
“Valgus extension overload test”	57
“Posteromedial impingement test elbow”	265
“Arm bar test posteromedial impingement”	15
“Medial epicondylitis test”	355
“Valgus stress test elbow”	2790
“Moving valgus stress test elbow”	632
“Milking Maneuver test elbow”	70
“Mill’s test elbow”	5630
“Maudsley’s test elbow”	309
“Cozen’s test elbow”	1870
“Grip and grind test elbow”	728
“Stand up test elbow”	10600
“Chair push-up test elbow”	3220
“Table-top relocation test elbow”	100
“Drawer sign elbow”	1630
“Lateral pivot shift test elbow”	1200
“Elbow exam” section on VuMedi	5
“Elbow anatomy and evaluation” section on Orthobullets	4
“Elbow” search on G6MD	8
Total	183802

Selection of videos

The first two-hundred videos on general physical examination and first fifty videos on specific elbow tests (sorted by relevance) on YouTube were included for initial screening. Because of the practically infinite output of the YouTube search engine, the authors decided on this arbitrary cut-off. All videos on overall and specific elbow examination on VuMedi, Orthobullets and G9MD were assessed. Videos had to be in English and in the format of an instructional video, i.e., videos in the form of a lecture (without moving video content), seminar, review, advertisement, and news or videos discussing history taking or symptoms were excluded. All videos were screened based on title and description on October 13, 2018. Videos were assessed and included in the study between October 13 and 31, 2018.

Data collection

For videos that met the inclusion criteria the following information was reported: title, duration of the video, URL, subject covered, days on YouTube, total number of views, amount of likes and dislikes, and the name, profession and type of uploader/creating organization. The uploader type was categorized as follows: university/school, hospital, informative website, private, business/company, other or unknown. For the profession of the uploader the following categories were used: doctor (e.g., orthopaedic surgeon, sports physician, general practitioner or rheumatologist), physiotherapist, student, other or unknown.

Accuracy and quality assessment

Assessment of the accuracy and quality of the included videos was performed by two independent raters (ELZ and RJM) using a modified scoring tool based on previous studies in this field with excellent inter-rater reliability [11, 13, 14, 31]. The scoring system is presented in **[figure 1].** Our modification aimed to increase focus on video accuracy instead of quality, including assessment of presentation of information on diagnostic accuracy. To rate the accuracy of the videos, a description of each test was provided to the raters as it was originally described in the literature[159]. A total of 18 points could be obtained: eight for the accuracy assessment and ten for the quality assessment. Half points were not allowed. Both categories contained major (two points each) and minor (one point each) criteria. Videos that scored the maximum amount of points amongst the major criteria by both observers were defined as educational useful videos. To test the reliability of the modified scoring system, we performed reliability analysis for intra- and inter-rater reliability on the first 20 videos before proceeding with the full sample.

Figure 1 – Scoring system

Accuracy assessment (total 8 points)

- Major criteria (2 points each):
  1. Arm position
  2. Examiner action
  3. Interpretation
- Minor criteria (1 point each):
  1. Verbal description
  2. Information on diagnostic accuracy

Quality assessment (total 10 points)

- Major criteria (2 points each):
  1. The video uses (simulated) patients to demonstrate the examination
  2. Images are clear
  3. Sounds are clear
  4. Creator/organisation providing the video are mentioned
- Minor criteria (1 point each):
  1. The video covers the topic identified in the title
  2. Designed at a level of undergraduate medical students

Statistical analysis

Data analysis was performed using Statistical Package for the Social Sciences 23 (IBM Corporation, Armonk, NY, USA). P-values <0.05 were considered statistically significant. Descriptive statistics and paired T-tests were used to examine demographics, accuracy, and quality of videos. 95% Confidence intervals (CI) were calculated using the group standard deviation in the total group and educationally non-useful videos. Due to the smaller sample size (<30), we used the total populations' standard deviation to calculate 95% CI for the educationally useful videos. Correlation between the accuracy and quality assessment was determined using Spearman's rho correlation coefficient. A correlation coefficient between 1.00 and 0.90 was considered very high, 0.90-0.70 high, 0.50-0.70 moderate, 0.50-0.30 low and 0.30-0.00 negligible, both in positive and negative direction[107]. Pearson's correlation coefficient was used to assess the correlation between the numbers of days on YouTube, views, likes and dislikes using the same cut-off values for interpretation. Independent T-tests and contingency tables were used to evaluate the influence of uploader type and profession on scores and educational usefulness.

To evaluate the reliability of the scoring system, we calculated the intra-rater reliability of rater one (EZ) and the inter-rater reliability for rater one and two (EZ and RM) using the intra-class correlation coefficient (ICC) with a two-way random effects model for the first 20 videos. Inter-rater reliability was considered excellent for ICC values between 1.00 and 0.75, good for values between 0.74 and 0.60, fair for values between 0.59 and 0.40, and poor for values less than 0.40[34]. It was not possible to calculate internal consistence using Cronbach's Alpha, due to the different scales for minor and major criteria and the small range (0-2 points and 0-1 points, respectively).

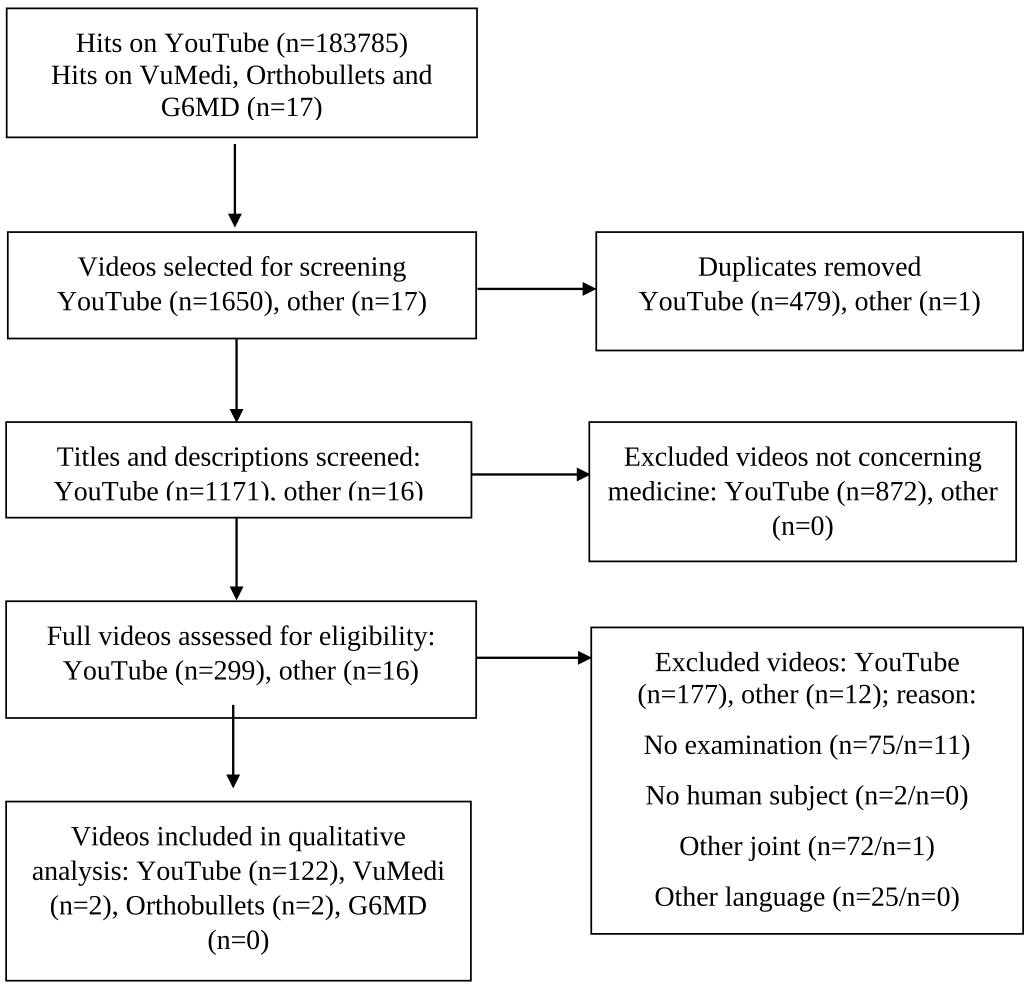
Finally, we calculated the inter-rater reliability of the scoring system for the complete dataset. Mean differences (MD) and the ICC with 95% CI were calculated using a two-way random effects model.

Results

Search

Our search resulted in 183,802 initial hits. One-thousand six-hundred and fifty video titles on YouTube were screened and 17 video titles were screened on VuMedi, Orthobullets, and G9MD. Of all screened videos, 7.4% (122/1,650) of the YouTube videos and 24% (4/17) of the videos from VuMedi, Orthobullets, and G9MD were eligible, resulting in a total of 126 videos (122/126 from YouTube, 96.8%) included in our study **[figure 2].**

Figure 2 – Flow chart of video selection



Characteristics of included videos

The 126 included videos were uploaded between June 2007 and February 2018. Video statistics for the total group and sorted by educational usefulness with mean, range and 95% CI are summarized in **[table 2].** Because of small number, dislikes per day online were not analysed.

Table 2 – Characteristics of included videos

	Educational useful (n=23)	Educational not useful (n=103)	Total (n=126)	p-value
Days online (mean (95%CI))	1017.04 (842 to 1190)	1935.23 (1750 to 2120)	1624.25 (1450 to 1800)	p<0.05

Views total (mean (95%CI))	6168.30 (1650 to 10700)	20211.94 (14600 to 25800)	16954.79 (12400 to 21500)	p<0.05
Views per day (mean (95% CI))	5.69 (2.14 to 7.86)	8.47 (5.88 to 10.1)	7.96 (5.25 to 8.75)	p<0.05
Likes (mean (95%CI))	32.35 (21.1 to 43.7)	32.56 (19.3 to 45.9)	31.44 (20.0 to 42.8)	p=0.989
Likes per day (mean (95% CI)	0.03 (0 to 0)	0.02 (0 to 0)	0.02 (0 to 0)	p=0.218
Dislikes (mean (95%CI))	0.30 (-0.22 to 0.82)	1.99 (1.3 to 2.68)	1.68 (0.48 to 1.52)	p=0.27

The uploader type varied from videos from informative websites (mostly by physiotherapists) (34%), private individuals without a website (27%), universities or university hospitals (22%), general or private hospitals (9%), and commercial companies (8%). For more than half of the videos the profession of the uploader was unknown (56%), with the other half of videos predominantly produced by physiotherapists (21%) and medical doctors (17%).

The content distribution of the included videos is summarized in **[table 3].** In 119 out of 126 videos (94%) a specific test was performed, with or without additional general examination. Stability tests of the medial collateral ligament (MCL) (n=46, 39% of the videos with a specific test) and lateral collateral ligament (LCL) (n=37, 31%), and specific tests for lateral epicondylitis (n=47, 39%) and medial epicondylitis (n=24, 20%) were covered most frequently.

Table 3 – Content distribution of included videos

Content	*Videos (n)*	*Percentage*
History taking	6	4.8%
Inspection	22	17.5%

Anatomy	15	11.9%
Carrying angle	10	7.9%
Range of motion	23	18.3%
Specific test	119	94.4%
Biceps	15	11.9%
Triceps	3	2.4%
VEOS	7	5.6%
ME	24	19.0%
MCL	46	36.5%
LE	47	37.3%
LCL	37	29.4%
OCD	2	1.6%

Pilot: reliability of the scoring system

Intra-rater reliability analysis showed excellent ICC's of the modified scoring system for accuracy (0.97; 95% CI 0.93-0.99), quality (0.97; 95% CI 0.93-0.99) and overall score (0.98; 95% CI 0.94-0.99). The inter-rater reliability analysis of the pilot sample showed excellent ICC's for accuracy (0.87; 95% CI 0.70-0.95), quality (0.92; 95% CI 0.79-0.97) and overall score (0.94; 95% CI 0.86-0.98) as well. Therefore, we further analysed the full sample of 126 videos with the scoring system as described in the methods section.

Accuracy and quality assessment

The mean accuracy assessment score of the total sample was 5.6/8 points (95% CI 5.3 to 5.9), ranging from 0 to 8 points. For videos on VuMedi, Orthobullets and G9MD, the mean accuracy assessment score was significantly higher (mean 6.5; 95% CI 6.2 to 6.8) compared to videos on YouTube (mean 5.6; 95% CI 5.3 to 5.9) ($P<0.001$).

Out of the maximum 10 points, the mean score for quality assessment was 7.3 points (95% CI 7.0 to 7.7), with a range from 4 to 10 points. For videos posted on the specific platforms, the mean quality assessment score was significantly higher (mean 10; 95% CI 9.6 to 10.0) compared to videos available on YouTube (mean 7.2; 95% CI 6.9 to 7.6) ($P<0.001$).

The mean overall score was 12.9 points (95% CI 12.4 to 13.5) out of a maximum of 18 points, ranging from 5 to 18 points. Mean overall score for videos on specific platforms was 16.5 (95% CI 16.0 to 17.0) and for videos on YouTube 12.8 (95% CI 12.3 to 13.3) ($P<0.001$).

In total, two out of 126 videos (1.6%) achieved the maximum accuracy score (8 points) by both observers (both available on YouTube). Thirty-one videos (24.6%) had the maximum score (10 points) on quality assessment, including all the educational videos from the specific platforms. In total, six out of 126 (4.8%) videos were given a maximum score (18 points) by one of the observers (all on YouTube), but in only two videos both observers agreed on the maximum score. Twenty-three videos (18.3%) fulfilled all major criteria by both observers and were therefore determined to be educationally useful **[table 4],** this includes three out of the four included videos posted on specific platforms.

Table 4 – Included videos and educational usefulness

Title	Duration	Link	Content	Mean overal score
Orthopaedics Video 6 - examination of the Elbow	06:57	https://www.youtube.com/watch?v=qJRCP67NG9c	General, medial/lateral epicondylitis and LCL stability tests	17
Wrist and Elbow Examination	11:26	https://www.youtube.com/watch?v=V6tgYwatAYU	General, medial/lateral epicondylitis, LCL/MCL stability tests	17
Elbow Exam Tests - Sugar Land Houston - Dr. J. Michael Benett	09:28	https://www.youtube.com/watch?v=2IoopfRaR6o	General, biceps/triceps, medial/lateral epicondylitis, LCL/MCL stability tests	17
The Reumatological examination of Elbows	02:18	https://www.youtube.com/watch?v=fYoeX5ZpC2w	General	17
Elbow exam	14:59	https://www.youtube.com/watch?v=Q-ef6inJ95E	General, biceps, medial/lateral epicondylitis, LCL/MCL stability tests	17,5
Musculoskeletal Examination and Joint injection Series_Examination of the Elbow	03:38	https://www.youtube.com/watch?v=LweM-dl6P2E	General	17
Physical Examination of the Elbow	11:43	https://www.orthobullets.com/video/view.aspx?id=1948	General, biceps/triceps, medial/lateral epicondylitis, LCL/MCL stability tests	17
Elbow exam	13:33	https://www.vumedi.com/video/elbow-exam/	General, medial/lateral epicondylitis, LCL/MCL stability tests	17
Physical Examination of the Elbow	10:28	https://www.vumedi.com/video/physical-examination-of-the-elbow/	General, biceps, medial/lateral epicondylitis, LCL/MCL stability tests	17
Biceps Squeeze Test	00:58	https://www.youtube.com/watch?v=HbOwTtbvrWM	Biceps test	17
LTA Diagnostic Tool: Biceps interval crease test	01:32	https://www.youtube.com/watch?v=OKL0PW5FD74	Biceps test	18
The valgus extension overload test	01:38	https://www.youtube.com/watch?v=HYG2spWs1qs	VEOS test	17,5
Elbow Valgus instability Stress test | Medial Collateral ligament	01:28	https://www.youtube.com/watch?v=3xF9_5fbJ8A	MCL stability test	17,5
The moving valgus stress test for MCL tears of the Elbow	02:08	https://www.youtube.com/watch?v=JIU_kv5VoQk	MCL stability test	18
The posterolateral rotatory drawer test for elbow instability	01:38	https://www.youtube.com/watch?v=y_Lm78EHvkM	LCL stability test	17
The Lateral Pivot Shift Apprehension Test | Posterolateral Rotatory Instability of the Elbow	01:26	https://www.youtube.com/watch?v=A3zExo7cmSc	LCL stability test	17
Elbow Varus Instability Stress test | Lateral Collateral Ligament	01:21	https://www.youtube.com/watch?v=5zl8GsG3hR4	LCL stability test	17
Moving Valgus stress test	01:03	https://www.youtube.com/watch?v=vI9LPr1w-00	MCL stability test	17
Moving Valgus Stress test	01:44	https://www.youtube.com/watch?v=4KeoOJb3864	MCL stability test	16
Elbow Valgus Stress Test	01:07	https://www.youtube.com/watch?v=3d2H_BAMsCE	MCL stability test	17
Lateral Epicondylitis Tests	00:30	https://www.youtube.com/watch?v=xa4op1Hv-L8	Lateral epicondylitis test	17
Push-Up Sign	00:49	https://www.youtube.com/watch?v=3KZqQCQJ9ac	LCL stability test	17
Chair sign	00:51	https://www.youtube.com/watch?v=azxNAY6Sr-w	LCL stability test	17,5

Correlation between the accuracy and quality assessment scores was considered weak, based on a Spearman's rho of 0.26 (P=0.003). The number of days online showed a weak negative correlation with the quality assessment (Pearson's rho of -0.26, P=0.003), and no correlation was found between the number of days online and video accuracy and overall score (p=0.090 and p=0.175). The total amount of views did not correlate with assessment, quality, or overall scores (p=0.938, p=0.674 and p=0.878 respectively). The number of likes showed a weak correlation with the quality score (r = 0.29, p<0.001) and overall score (r = 0.2, p=0.026), but not with the accuracy of the video (r = -0.07, p=0.941). The total number of dislikes was too small (n = 212) to draw conclusions on correlations.

Educational usefulness

The mean number of days online and viewers was significantly lower in educational useful videos compared to less useful educational videos (p<0.01; p=0.02). No significant difference was observed for the amount of likes between these two groups (p=0.99). There was a significant association between the type of uploader and usefulness of video content (p<0.001). Most educationally useful videos, originated from websites, universities/ university hospitals, and general hospitals (48%, 26%, and 26% respectively). None of the videos uploaded by private persons, companies or other/unknown uploaders were classified as educational useful. Within the uploader type, general hospitals and websites had the highest scores (55% and 26% of videos were useful respectively, **[figure 3]).** Additional analysis of all included videos showed no significant difference for accuracy scores in comparison to content from websites, universities/university hospitals and general hospitals (p=0.534), but around 0.5 to 1 point difference in quality and overall scores (p<0.001).

There was no significant impact of uploaders' profession on educational usefulness (p=0.110). Notable differences were observed between unknown professions (only 11% of videos in this group was useful) and medical doctors, physiotherapists and students (38%, 19% and 20% respectively**) [figure 4].** Comparison of videos developed by unknown professions with videos from known professions showed a significant difference in quality and overall scores (p<0.001), but not in accuracy scores (p=0.220). The mean quality score was 8.1 for videos from known professions and 6.7 for videos from unknown professions, with mean overall scores of 13.9 and 12.2, respectively.

Figure 3 – Bar chart showing influence of uploader type on educational usefulness

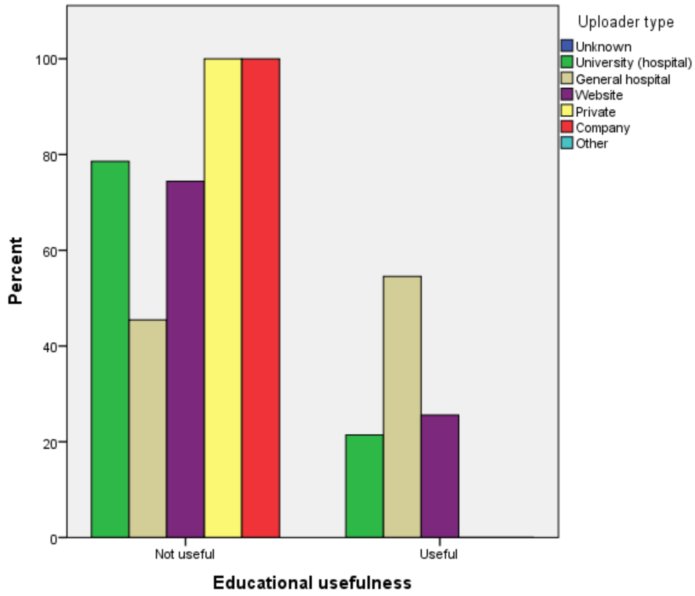
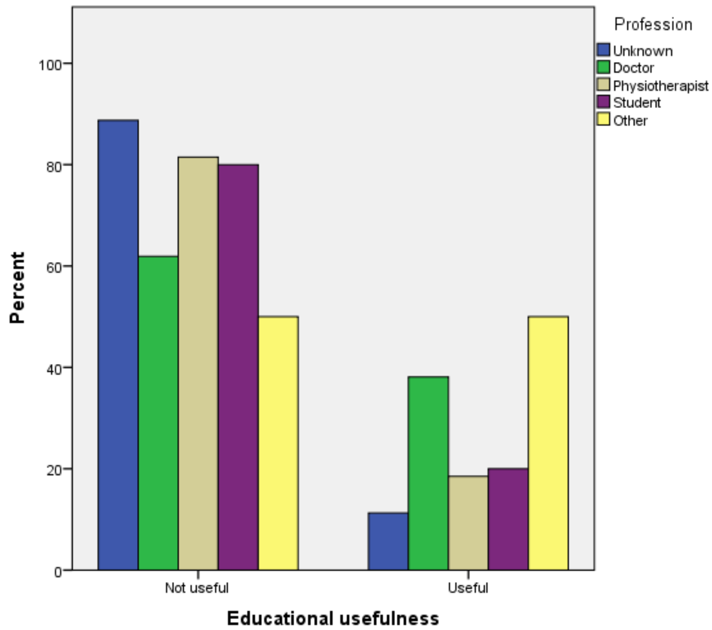


Figure 4 – Bar chart showing influence of profession of uploader on educational usefulness



Inter-rater reliability

For the accuracy and quality assessment, the mean difference between the two observers was significant (p<0.001), however, mean differences were less than 0.5 point. The mean difference for the overall score between rater one and rater two was not significant, with a mean difference of 0.024 (p=0.871).

Inter-rater reliability analysis for the accuracy assessment showed an excellent ICC of 0.80 (95% CI 0.70-0.86). Quality assessment also showed an excellent reliability with an ICC of 0.91 (95% CI 0.85-0.95). ICC's of the overall score was 0.93 (95% CI 0.09-0.95).

Discussion

This study assessed the accuracy and quality of currently available educational videos on physical examination of the elbow. The aim of this study was to provide characteristics of accurate and qualitative videos and a list of currently available educationally valuable videos. The findings of this study may guide students and residents in identifying educationally useful videos and to help health care providers with the development of future educational online content.

Main observations

A large variability of scores for both accuracy and quality were observed for the 126 videos reviewed in this study. Accuracy and quality assessment had a negligible correlation, suggesting that high-quality videos are not necessarily accurate in their presentation of physical examination and vice versa. Out of the included studies, 23 fulfilled all major criteria and were therefore classified as educationally useful. Notably, three out of the four included videos from specialized platforms (VuMedi and Orthobullets) were rated as educationally valuable and all showed higher accuracy, quality and overall scores compared to YouTube videos. Given the difficulty of regulating content on YouTube, these results are not surprising. This finding confirms our hypothesis that specialized platforms provide videos of higher quality and accuracy than videos from YouTube. Only two out of 126 videos achieved a maximum score by both observers. Overall, more videos received the maximum score for quality assessment than for accuracy assessment. One of the reasons for this is that only a few videos addressed diagnostic accuracy of the tests described. In previously described scoring systems, addressing the diagnostic accuracy was not a part of the accuracy assessment [11, 13, 14]. However, to interpret the results of physical examination tests, knowledge on the diagnostic accuracy is of vital importance to rule in or out a diagnosis by physical examination. Notably, all examination videos on specialized platforms lacked information on the diagnostic accuracy. Interrater reliability of the modified scoring system showed excellent ICCs, especially when accuracy and quality assessment were combined. Mean scores for the combined accuracy and quality scores showed no significant difference.

With the abundance of online available content and increasing focus on self-learning, it is important for students and residents to identify educationally high quality videos. We observed that the majority of educational useful videos on elbow physical examination were provided by general hospitals and websites from physiotherapy practices or private individuals and to a lesser extent by universities and university hospitals. The limited contribution of university hospitals (as education institutions) may be explained by the fact that extensive physical examination of the elbow generally extends beyond the content of the general medical study curriculum. Between general hospitals and websites only quality scores differed (approximately 1 point higher score for websites compared to general hospitals), which may be explained by sponsorship or profit motives of more commercially oriented companies/websites. Considering the profession of the uploader, our study show that when the profession is unknown, the quality and overall score of educational videos are lower compared to clips of known professions. Therefore, we recommend videos from providers of which the profession is presented for educational purposes. Comparing our findings with observations in previous literature on medical examination videos, only one study from Urch et al. (2016) compared videos about shoulder examination on YouTube with specialised platforms. Their conclusion supported our conclusion: videos on YouTube were less accurate compared to specialised platforms (VuMedi, G6MD and orthobullets)[151]. All other studies available only included YouTube videos[11, 13, 14, 21, 83]. Their results were mostly in accordance with our study. For example, Lee et al (2018) investigated the physical examination of the shoulder and found that videos originating from (known) physicians are more useful comparing to videos from (unknown) individuals[83]. Also the limitations of the YouTube search algorithm and video regulations leading to a relative large number of screened videos and small number of educational useful videos were addressed in most studies. The studies by Azer (2012, 2012 and 2013) on the nervous system, surface anatomy and cardiorespiratory system used the scoring system we based our modification on and showed comparable scores for educational useful and non-useful videos[11, 13, 14].

Strengths and limitations

Because YouTube content changes continuously, potentially useful videos may be missed. However, search results often not match with search terms: our search yielded almost 200.000 hits, of which many did not focus on physical examination, medicine, and/or the elbow joint. Therefore we pragmatically decided to only screen titles of the first 1,650 videos. This problem is not present on video platforms that target medical professionals, however these platforms are often not known by non-specialized care providers such as medical students, general practitioners and physiotherapists and provide less videos. Improvement of YouTube's search algorithm system may lead to more accurate matching of search terms and resulting video content.

The inclusion of videos for this study was limited to videos in English, so that the information is understandable for most viewers. These factors might have led to a selection bias and a limited amount of educationally useful videos. Furthermore, the scoring system used in this study was not validated before, but limited modifications were made to the previously validated scoring system and our analysis showed excellent inter-rater reliability.

Recommendations and future directions

Teachers and clinical supervisors should be aware that students and residents use open-access online learning platforms such as YouTube and recognize its pitfalls. Our study shows that numerous teachers and clinicians are creating online content, but that these videos frequently lack accuracy and/or quality. We advise content creators to film in a quiet room with good quality electronic video capture systems. Creators should provide step-by-step physical examination with clear verbal instructions and images. Furthermore, we advise the examiners to introduce themselves by providing information on their profession and institution. In order to enable the viewer to interpret the physical examination properly, information on diagnostic accuracy should be provided when available; in cases where there are no diagnostic accuracy studies available this should be indicated as well. Students and residents should be aware that not all online content is accurate and qualitative and that videos posted on specialized platforms such as VuMedi and Orthobullets are generally more educational useful compared to videos on YouTube. Viewership, likes/dislikes, and days online are not appropriate to determine whether or not a video is educationally useful.

For future research, it would be interesting to investigate the educational value of videos for students and residents in comparison to written content with or without static images and live education by a teacher or clinician. Furthermore, it may be useful to develop creator guidelines for physical examination videos for educational purpose such as the LAP-VEGaS guidelines for videos on laparoscopic cholecystectomy[31]. The modified assessment tool used in this study might serve for this goal.

Conclusion

In this study, we indicated 23 educational useful videos on the physical examination of the elbow. Videos posted on specialized platforms, such as VuMedi and Orthobullets, by a creator of which the profession is known are generally more reliable compared to videos of which the creator's background and/or institution is unknown. Viewership, likes/dislikes, and days online do not indicate usefulness. The assessment tool used in this study for evaluating accuracy and quality of videos is easy to apply and covers key elements of good-quality educational videos. The tool can be used by students and residents to assess reliability of educational video content and aid educators in the development of new online content.

References

References apply to all articles in this paper and are numbered continuously.

1. Physical status: the use and interpretation of anthropometry. Report of a WHO Expert Committee. World Health Organ Tech Rep Ser 1995;854:1-452.
2. Akgun T, Karabay CY, Kocabay G, Kalayci A, Oduncu V, Guler A, et al. Learning electrocardiogram on YouTube: how useful is it? J Electrocardiol 2014;47(1):113-7. doi:10.1016/j.jelectrocard.2013.09.004
3. Aldridge JW, Bruno RJ, Strauch RJ, Rosenwasser MP. Management of acute and chronic biceps tendon rupture. Hand Clin 2000;16(3):497-503.
4. Anakwenze OA, Kancherla VK, Iyengar J, Ahmad CS, Levine WN. Posterolateral rotatory instability of the elbow. Am J Sports Med 2014;42(2):485-91. doi:0363546513494579 [pii];10.1177/0363546513494579 [doi]
5. Armstrong AD, MacDermid JC, Chinchalkar S, Stevens RS, King GJ. Reliability of range-of-motion measurement in the elbow and forearm. J Shoulder Elbow Surg 1998;7(6):573-80.
6. Aronson ID, Plass JL, Bania TC. Optimizing educational video through comparative trials in clinical environments. Educ Technol Res Dev 2012;60(3):469-82.
7. Arrigoni P, Cucchi D, D'Ambrosi R, Butt U, Safran MR, Denard P, et al. Intra-articular findings in symptomatic minor instability of the lateral elbow (SMILE). Knee Surg Sports Traumatol Arthrosc 2017;25(7):2255-63. doi:10.1007/s00167-017-4530-x
8. Arrigoni P, Cucchi D, Menon A, Randelli P. It's time to change perspective! New diagnostic tools for lateral elbow pain. Musculoskelet Surg 2017;101(Suppl 2):175-9. doi:10.1007/s12306-017-0486-8
9. Arvind CH, Hargreaves DG. Tabletop relocation test: a new clinical test for posterolateral rotatory instability of the elbow. J Shoulder Elbow Surg 2006;15(6):707-8. doi:S1058-2746(06)00075-9 [pii];10.1016/j.jse.2006.01.005 [doi]
10. Ashida T, Sugiyama T, Okuno S, Ebihara A, Fujii J. Relationship between home blood pressure measurement and medication compliance and name recognition of antihypertensive drugs. Hypertens Res 2000;23(1):21-4.
11. Azer SA. Can "YouTube" help students in learning surface anatomy? Surg Radiol Anat 2012;34(5):465-8. doi:10.1007/s00276-012-0935-x
12. Azer SA. Understanding pharmacokinetics: are YouTube videos a useful learning resource? Eur Rev Med Pharmacol Sci 2014;18(13):1957-67.

13. Azer SA, Aleshaiwi SM, Algrain HA, Alkhelaif RA. Nervous system examination on YouTube. BMC Med Educ 2012;12:126. doi:10.1186/1472-6920-12-126
14. Azer SA, Algrain HA, AlKhelaif RA, AlEshaiwi SM. Evaluation of the educational value of YouTube videos about physical examination of the cardiovascular and respiratory systems. J Med Internet Res 2013;15(11):e241. doi:10.2196/jmir.2728
15. Barrett E, McCreesh K, Lewis J. Reliability and validity of non-radiographic methods of thoracic kyphosis measurement: a systematic review. Man Ther 2014;19(1):10-7. doi:10.1016/j.math.2013.09.003
16. Bisson L, Moyer M, Lanighan K, Marzo J. Complications associated with repair of a distal biceps rupture using the modified two-incision technique. J Shoulder Elbow Surg 2008;17(1 Suppl):67S-71S. doi:10.1016/j.jse.2007.04.008
17. Bland JM. Introduction to medical statistics. Oxford: Oxford University Press; 2006.
18. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1(8476):307-10.
19. Blonna D, Zarkadas PC, Fitzsimmons JS, O'Driscoll SW. Accuracy and inter-observer reliability of visual estimation compared to clinical goniometry of the elbow. Knee Surg Sports Traumatol Arthrosc 2012;20(7):1378-85. doi:10.1007/s00167-011-1720-9
20. Blonna D, Zarkadas PC, Fitzsimmons JS, O'Driscoll SW. Validation of a photography-based goniometry method for measuring joint range of motion. J Shoulder Elbow Surg 2012;21(1):29-35. doi:10.1016/j.jse.2011.06.018
21. Borgersen NJ, Henriksen MJ, Konge L, Sorensen TL, Thomsen AS, Subhi Y. Direct ophthalmoscopy on YouTube: analysis of instructional YouTube videos' content and approach to visualization. Clin Ophthalmol 2016;10:1535-41. doi:10.2147/OPTH.S111648
22. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. BMJ 2015;351:h5527. doi:10.1136/bmj.h5527
23. Boulos MN, Maramba I, Wheeler S. Wikis, blogs and podcasts: a new generation of Web-based tools for virtual collaborative clinical practice and education. BMC Med Educ 2006;6:41. doi:10.1186/1472-6920-6-41
24. Bray EP, Holder R, Mant J, McManus RJ. Does self-monitoring reduce blood pressure? Meta-analysis with meta-regression of randomized controlled trials. Ann Med 2010;42(5):371-86. doi:10.3109/07853890.2010.489567
25. Brosseau L, Tousignant M, Budd J, Chartier N, Duciaume L, Plamondon S, et al. Intratester and intertester reliability and criterion validity of the parallelogram and universal goniometers for active knee flexion in healthy subjects. Physiother Res Int 1997;2(3):150-66.

26. Burton A. YouTube-ing your way to neurological knowledge. Lancet Neurol 2008;7(12):1086-7. doi:10.1016/S1474-4422(08)70252-7
27. C.Norkin C, D.White J. Measurement of joint motion: a guide to goniometry.; 2003.
28. Caekebeke P, Schenkels E, Bell SN, van Riet R. Distal Biceps Provocation Test. J Hand Surg Am 2021;46(8):710 e1- e4. doi:10.1016/j.jhsa.2020.12.012
29. Cardoso AF, Moreli L, Braga FT, Vasques CI, Santos CB, Carvalho EC. Effect of a video on developing skills in undergraduate nursing students for the management of totally implantable central venous access ports. Nurse Educ Today 2012;32(6):709-13. doi:10.1016/j.nedt.2011.09.012
30. Casado MI, Castano G, Arraez-Aybar LA. Audiovisual material as educational innovation strategy to reduce anxiety response in students of human anatomy. Adv Health Sci Educ Theory Pract 2012;17(3):431-40. doi:10.1007/s10459-011-9307-2
31. Celentano V, Smart N, McGrath J, Cahill RA, Spinelli A, Obermair A, et al. LAP-VEGaS Practice Guidelines for Reporting of Educational Videos in Laparoscopic Surgery: A Joint Trainers and Trainees Consensus Statement. Ann Surg 2018;268(6):920-6. doi:10.1097/SLA.0000000000002725
32. Chapleau J, Canet F, Petit Y, Laflamme GY, Rouleau DM. Validity of goniometric elbow measurements: comparative study with a radiographic method. Clin Orthop Relat Res 2011;469(11):3134-40. doi:10.1007/s11999-011-1986-8 [doi]
33. Chen FS, Diaz VA, Loebenberg M, Rosen JE. Shoulder and elbow injuries in the skeletally immature athlete. J Am Acad Orthop Surg 2005;13(3):172-85.
34. Cicchetti DV. Multiple comparison methods: establishing guidelines for their valid application in neuropsychological research. J Clin Exp Neuropsychol 1994;16(1):155-61. doi:10.1080/01688639408402625
35. Cimatti B, Marcolino AM, Barbosa RI, De Cassia Registro Fonseca M. A study to compare two goniometric methods for measuring active pronation and supination range of motion. Hand Therapy 2013:57-63.
36. Cleffken B, van BG, van MH, Brink P, Olde DS. Test-retest reproducibility of elbow goniometric measurements in a rigid double-blinded protocol: intervals for distinguishing between measurement error and clinical change. J Shoulder Elbow Surg 2007;16(6):788-94. doi:S1058-2746(07)00360-6 [pii];10.1016/j.jse.2007.02.134 [doi]
37. Cook C, Hegedus E. Orthopedic Physical Examination Tests. Pearson Education; 2012.
38. D.Bracker M. The 5-Minute Sports Medicine Consult. 2nd ed.: Philadelphia USA: Wolters Kluwer Lippincott Williams & Wilkins; 2011.

39. D.Miller M, R.Thompson S. DeLee & Drez's Orthopaedic Sports Medicine. 4 ed.: Saunders Elsevier Inc.; 2014.

40. de la Fuente J, Blasi M, Martinez S, Barcelo P, Cachan C, Miguel M, et al. Ultrasound classification of traumatic distal biceps brachii tendon injuries. Skeletal Radiol 2018;47(4):519-32. doi:10.1007/s00256-017-2816-1

41. Devereaux MW, ElMaraghy AW. Improving the rapid and reliable diagnosis of complete distal biceps tendon rupture: a nuanced approach to the clinical examination. Am J Sports Med 2013;41(9):1998-2004. doi:0363546513493383 [pii];10.1177/0363546513493383 [doi]

42. Docherty MA, Schwab RA, Ma OJ. Can elbow extension be used as a test of clinically significant injury? South Med J 2002;95(5):539-41.

43. Dorf ER, Chhabra AB, Golish SR, McGinty JL, Pannunzio ME. Effect of elbow position on grip strength in the evaluation of lateral epicondylitis. J Hand Surg Am 2007;32(6):882-6. doi:S0363-5023(07)00436-4 [pii];10.1016/j.jhsa.2007.04.010 [doi]

44. Dugas JR. Valgus extension overload: diagnosis and treatment. Clin Sports Med 2010;29(4):645-54. doi:S0278-5919(10)00051-7 [pii];10.1016/j.csm.2010.07.001 [doi]

45. Eddy DM. Clinical decision making: from theory to practice. Cost-effectiveness analysis. Will it be accepted? JAMA 1992;268(1):132-6.

46. ElMaraghy A, Devereaux M. The "bicipital aponeurosis flex test": evaluating the integrity of the bicipital aponeurosis and its implications for treatment of distal biceps tendon ruptures. J Shoulder Elbow Surg 2013;22(7):908-14. doi:S1058-2746(13)00130-4 [pii];10.1016/j.jse.2013.02.005 [doi]

47. ElMaraghy A, Devereaux M, Tsoi K. The biceps crease interval for diagnosing complete distal biceps tendon ruptures. Clin Orthop Relat Res 2008;466(9):2255-62. doi:10.1007/s11999-008-0334-0 [doi]

48. Ferriero G, Vercelli S, Sartorio F, Munoz Lasa S, Ilieva E, Brigatti E, et al. Reliability of a smartphone-based goniometer for knee joint goniometry. Int J Rehabil Res 2013;36(2):146-51. doi:10.1097/MRR.0b013e32835b8269

49. Festa A, Mulieri PJ, Newman JS, Spitz DJ, Leslie BM. Effectiveness of magnetic resonance imaging in detecting partial and complete distal biceps tendon rupture. J Hand Surg Am 2010;35(1):77-83. doi:10.1016/j.jhsa.2009.08.016

50. Fieseler G, Molitor T, Irlenbusch L, Delank KS, Laudner KG, Hermassi S, et al. Intrarater reliability of goniometry and hand-held dynamometry for shoulder and elbow examinations in female team handball athletes and asymptomatic volunteers. Arch Orthop Trauma Surg 2015. doi:10.1007/s00402-015-2331-6 [doi];10.1007/s00402-015-2331-6 [pii]

51. Fincher RM, Simpson DE, Mennin SP, Rosenfeld GC, Rothman A, McGrew MC, et al. Scholarship in teaching: an imperative for the 21st century. Acad Med 2000;75(9):887-94.
52. Fish DR, Wingate L. Sources of goniometric error at the elbow. Phys Ther 1985;65(11):1666-70.
53. Fleisig GS, Andrews JR, Dillman CJ, Escamilla RF. Kinetics of baseball pitching with implications about injury mechanisms. Am J Sports Med 1995;23(2):233-9.
54. Fleiss JL. Analysis of data from multiclinic trials. Control Clin Trials 1986;7(4):267-75.
55. Flowers KR, Stephens-Chisar J, LaStayo P, Galante BL. Intrarater reliability of a new method and instrumentation for measuring passive supination and pronation: a preliminary study. J Hand Ther 2001;14(1):30-5.
56. Gajdosik RL. Comparison and reliability of three goniometric methods for measuring forearm supination and pronation. Percept Mot Skills 2001;93(2):353-5. doi:10.2466/pms.2001.93.2.353 [doi]
57. Gajdosik RL, Bohannon RW. Clinical measurement of range of motion. Review of goniometry emphasizing reliability and validity. Phys Ther 1987;67(12):1867-72.
58. Golden DW, Jhee JT, Gilpin SP, Sawyer JR. Elbow range of motion and clinical carrying angle in a healthy pediatric population. J Pediatr Orthop B 2007;16(2):144-9. doi:10.1097/BPB.0b013e328010fcf2 [doi];01202412-200703000-00013 [pii]
59. Golden DW, Wojcicki JM, Jhee JT, Gilpin SL, Sawyer JR, Heyman MB. Body mass index and elbow range of motion in a healthy pediatric population: a possible mechanism of overweight in children. J Pediatr Gastroenterol Nutr 2008;46(2):196-201. doi:10.1097/MPG.0b013e31812f568b [doi];00005176-200802000-00012 [pii]
60. Goodwin J, Clark C, Deakes J, Burdon D, Lawrence C. Clinical methods of goniometry: a comparative study. Disabil Rehabil 1992;14(1):10-5.
61. Gorgos KS, Wasylyk NT, Van Lunen BL, Hoch MC. Inter-clinician and intra-clinician reliability of force application during joint mobilization: a systematic review. Man Ther 2014;19(2):90-6. doi:10.1016/j.math.2013.12.003
62. Grimes DA, Schulz KF. Refining clinical diagnosis with likelihood ratios. Lancet 2005;365(9469):1500-5. doi:S0140-6736(05)66422-7 [pii];10.1016/S0140-6736(05)66422-7 [doi]
63. Gunal I, Kose N, Erdogan O, Gokturk E, Seber S. Normal range of motion of the joints of the upper extremity in male subjects, with special reference to side. J Bone Joint Surg Am 1996;78(9):1401-4.
64. Haidich AB. Meta-analysis in medical research. Hippokratia 2010;14(Suppl 1):29-37.

65. Hardaker WT, Jr., Garrett WE, Jr., Bassett FH, 3rd. Evaluation of acute traumatic hemarthrosis of the knee joint. South Med J 1990;83(6):640-4.
66. Harding WG, 3rd. A new clinical test for avulsion of the insertion of the biceps tendon. Orthopedics 2005;28(1):27-9.
67. Hausman MR, Lang P. Examination of the elbow: current concepts. J Hand Surg Am 2014;39(12):2534-41. doi:S0363-5023(14)00504-8 [pii];10.1016/j.jhsa.2014.04.028 [doi]
68. Hawi N, Liodakis E, Musolli D, Suero EM, Stuebig T, Claassen L, et al. Range of motion assessment of the shoulder and elbow joints using a motion sensing input device: a pilot study. Technol Health Care 2014;22(2):289-95. doi:2328658713651L71 [pii];10.3233/THC-140831 [doi]
69. Hegedus EJ, Cook C, Lewis J, Wright A, Park JY. Combining orthopedic special tests to improve diagnosis of shoulder pathology. Phys Ther Sport 2015;16(2):87-92. doi:10.1016/j.ptsp.2014.08.001
70. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003;327(7414):557-60. doi:10.1136/bmj.327.7414.557
71. Hoffmann T, Russell T, Cooke H. Remote measurement via the Internet of upper limb range of motion in people who have had a stroke. J Telemed Telecare 2007;13(8):401-5. doi:10.1258/135763307783064377
72. Jobe FW, Nuber G. Throwing injuries of the elbow. Clin Sports Med 1986;5(4):621-36.
73. Kane SF, Lynch JH, Taylor JC. Evaluation of elbow pain in adults. Am Fam Physician 2014;89(8):649-57. doi:d10823 [pii]
74. Karagiannopoulos C, Sitler M, Michlovitz S. Reliability of 2 functional goniometric methods for measuring forearm pronation and supination active range of motion. J Orthop Sports Phys Ther 2003;33(9):523-31. doi:10.2519/jospt.2003.33.9.523 [doi]
75. Keene AB, Shiloh AL, Dudaie R, Eisen LA, Savel RH. Online testing from Google Docs to enhance teaching of core topics in critical care: a pilot study. Med Teach 2012;34(12):1075-7. doi:10.3109/0142159X.2012.716553
76. Kelly EW, Steinmann S, O'Driscoll SW. Surgical treatment of partial distal biceps tendon ruptures through a single posterior incision. J Shoulder Elbow Surg 2003;12(5):456-61. doi:10.1016/s1058-2746(03)00052-1
77. Kelly MP, Perkinson SG, Ablove RH, Tueting JL. Distal Biceps Tendon Ruptures: An Epidemiological Analysis Using a Large Population Database. Am J Sports Med 2015;43(8):2012-7. doi:10.1177/0363546515587738

78. Kim SG, Kim EK. Test-retest reliability of an active range of motion test for the shoulder and hip joints by unskilled examiners using a manual goniometer. J Phys Ther Sci 2016;28(3):722-4. doi:10.1589/jpts.28.722

79. King JW, Brelsford HJ, Tullos HS. Analysis of the pitching arm of the professional baseball pitcher. Clin Orthop Relat Res 1969;67:116-23.

80. Kingsley K, Galbraith GM, Herring M, Stowers E, Stewart T, Kingsley KV. Why not just Google it? An assessment of information literacy skills in a biomedical science curriculum. BMC Med Educ 2011;11:17. doi:10.1186/1472-6920-11-17

81. Koppenhaven DS, Cleland J. Netter's Orthopaedic Clinical Examination. Elsevier; 2010.

82. Kotnis NA, Chiavaras MM, Harish S. Lateral epicondylitis and beyond: imaging of lateral elbow pain with clinical-radiologic correlation. Skeletal Radiol 2012;41(4):369-86. doi:10.1007/s00256-011-1343-8

83. Lee H, Choi A, Jang Y, Lee JI. YouTube as a learning tool for four shoulder tests. Prim Health Care Res Dev 2018:1-7. doi:10.1017/S1463423618000804

84. Lee JK, Yao L, Phelps CT, Wirth CR, Czajka J, Lozman J. Anterior cruciate ligament tears: MR imaging compared with arthroscopy and clinical tests. Radiology 1988;166(3):861-4. doi:10.1148/radiology.166.3.3340785

85. Lennon RI, Riyat MS, Hilliam R, Anathkrishnan G, Alderson G. Can a normal range of elbow movement predict a normal elbow x ray? Emerg Med J 2007;24(2):86-8. doi:24/2/86 [pii];10.1136/emj.2006.039792 [doi]

86. Levy HJ, Mashoof AA, Morgan D. Repair of chronic ruptures of the distal biceps tendon using flexor carpi radialis tendon graft. Am J Sports Med 2000;28(4):538-40. doi:10.1177/03635465000280041501

87. Lijmer JG, Mol BW, Heisterkamp S, Bonsel GJ, Prins MH, van der Meulen JH, et al. Empirical evidence of design-related bias in studies of diagnostic tests. JAMA 1999;282(11):1061-6. doi:joc90732 [pii]

88. Lipsey MW, Wilson D. Practical Meta-Analysis (Applied Social Research Methods). SAGE Publications; 2000.

89. Loong TW. Understanding sensitivity and specificity with the right side of the brain. BMJ 2003;327(7417):716-9. doi:10.1136/bmj.327.7417.716

90. Low JL. The reliability of joint measurement. Physiotherapy 1976;62(7):227-9.

91. Lucas N, Macaskill P, Irwig L, Moran R, Rickards L, Turner R, et al. The reliability of a quality appraisal tool for studies of diagnostic reliability (QAREL). BMC Med Res Methodol 2013;13:111. doi:10.1186/1471-2288-13-111

92. Luokkala T, Siddharthan SK, Karjalainen TV, Watts AC. Distal biceps hook test - Sensitivity in acute and chronic tears and ability to predict the need for graft reconstruction. Shoulder Elbow 2020;12(4):294-8. doi:10.1177/1758573219847146
93. Lynch J, Yu CC, Chen C, Muh S. Magnetic resonance imaging versus ultrasound in diagnosis of distal biceps tendon avulsion. Orthop Traumatol Surg Res 2019;105(5):861-6. doi:10.1016/j.otsr.2019.01.021
94. MacDermid JC, Michlovitz SL. Examination of the elbow: linking diagnosis, prognosis, and outcomes as a framework for maximizing therapy interventions. J Hand Ther 2006;19(2):82-97. doi:S0894-1130(06)00059-7 [pii];10.1197/j.jht.2006.02.018 [doi]
95. Macedo LG, Magee DJ. Differences in range of motion between dominant and nondominant sides of upper and lower extremities. J Manipulative Physiol Ther 2008;31(8):577-82. doi:S0161-4754(08)00237-6 [pii];10.1016/j.jmpt.2008.09.003 [doi]
96. Martens MJ, Duvivier RJ, van Dalen J, Verwijnen GM, Scherpbier AJ, van der Vleuten CP. Student views on the effective teaching of physical examination skills: a qualitative study. Med Educ 2009;43(2):184-91. doi:10.1111/j.1365-2923.2008.03283.x
97. May S, Chance-Larsen K, Littlewood C, Lomas D, Saad M. Reliability of physical examination tests used in the assessment of patients with shoulder problems: a systematic review. Physiotherapy 2010;96(3):179-90. doi:10.1016/j.physio.2009.12.002
98. Meislin MA, Wagner ER, Shin AY. A Comparison of Elbow Range of Motion Measurements: Smartphone-Based Digital Photography Versus Goniometric Measurements. J Hand Surg Am 2016;41(4):510-5 e1. doi:10.1016/j.jhsa.2016.01.006
99. Metzman LS, Tivener KA. The Supination-Pronation Test for Distal Biceps Tendon Rupture. Am J Orthop (Belle Mead NJ) 2015;44(10):E361-E4.
100. Milanese S, Gordon S, Buettner P, Flavell C, Ruston S, Coe D, et al. Reliability and concurrent validity of knee angle measurement: smart phone app versus universal goniometer used by experienced and novice clinicians. Man Ther 2014;19(6):569-74. doi:10.1016/j.math.2014.05.009
101. Miller MD. SMART! Sports Medicine Assessment and Review Textbook. Elsevier; 2010.
102. Miyamoto RG, Elser F, Millett PJ. Distal biceps tendon injuries. J Bone Joint Surg Am 2010;92(11):2128-38. doi:92/11/2128 [pii];10.2106/JBJS.I.01213 [doi]
103. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA Statement. Open Med 2009;3(3):e123-e30.
104. Morrey BF. Applied anatomy and biomechanics of the elbow joint. Instr Course Lect 1986;35:59-68.

105. Morrey BF, An KN. Articular and ligamentous contributions to the stability of the elbow joint. Am J Sports Med 1983;11(5):315-9.

106. Morrey BF, Askew LJ, Chao EY. A biomechanical study of normal functional elbow motion. J Bone Joint Surg Am 1981;63(6):872-7.

107. Mukaka MM. Statistics corner: A guide to appropriate use of correlation coefficient in medical research. Malawi Med J 2012;24(3):69-71.

108. Naylor JM, Ko V, Adie S, Gaskin C, Walker R, Harris IA, et al. Validity and reliability of using photography for measuring knee range of motion: a methodological study. BMC Musculoskelet Disord 2011;12:77. doi:10.1186/1471-2474-12-77

109. Nesterenko S, Domire ZJ, Morrey BF, Sanchez-Sotelo J. Elbow strength and endurance in patients with a ruptured distal biceps tendon. J Shoulder Elbow Surg 2010;19(2):184-9. doi:10.1016/j.jse.2009.06.001

110. Newcombe RG. Two-sided confidence intervals for the single proportion: comparison of seven methods. Stat Med 1998;17(8):857-72.

111. O'Driscoll SW, Goncalves LB, Dietz P. The hook test for distal biceps tendon avulsion. Am J Sports Med 2007;35(11):1865-9. doi:0363546507305016 [pii];10.1177/0363546507305016 [doi]

112. O'Driscoll SW, Lawton RL, Smith AM. The "moving valgus stress test" for medial collateral ligament tears of the elbow. Am J Sports Med 2005;33(2):231-9.

113. Ockendon M, Gilbert RE. Validation of a novel smartphone accelerometer-based knee goniometer. J Knee Surg 2012;25(4):341-5. doi:10.1055/s-0031-1299669

114. Park W, Ramachandran J, Weisman P, Jung ES. Obesity effect on male active joint range of motion. Ergonomics 2010;53(1):102-8. doi:918525720 [pii];10.1080/00140130903311617 [doi]

115. Petherick M, Rheault W, Kimble S, Lechner C, Senear V. Concurrent validity and intertester reliability of universal and fluid-based goniometers for active elbow range of motion. Phys Ther 1988;68(6):966-9.

116. Pewsner D, Battaglia M, Minder C, Marx A, Bucher HC, Egger M. Ruling a diagnosis in or out with "SpPIn" and "SnNOut": a note of caution. BMJ 2004;329(7459):209-13. doi:10.1136/bmj.329.7459.209

117. Philippou A, Maridaki M, Bogdanis GC. Angle-specific impairment of elbow flexors strength after isometric exercise at long muscle length. J Sports Sci 2003;21(10):859-65. doi:10.1080/0264041031000140356 [doi]

118. Pitts SR, Niska RW, Xu J, Burt CW. National Hospital Ambulatory Medical Care Survey: 2006 emergency department summary. Natl Health Stat Report 2008(7):1-38.

119. Polkinghorn BS. A novel method for assessing elbow pain resulting from epicondylitis. J Chiropr Med 2002;1(3):117-21. doi:S0899-3467(07)60015-9 [pii];10.1016/S0899-3467(07)60015-9 [doi]

120. Pomerance J. Radiographic analysis of lateral epicondylitis. J Shoulder Elbow Surg 2002;11(2):156-7.

121. Portney LG, Watkins MP. Foundations of clinical research: applications to practise.: Upper Saddle River: Prentice-Hall; 2000.

122. Rahman RK, Levine WN, Ahmad CS. Elbow medial collateral ligament injuries. Curr Rev Musculoskelet Med 2008;1(3-4):197-204. doi:10.1007/s12178-008-9026-3

123. Raikos A, Waidyasekara P. How useful is YouTube in learning heart anatomy? Anat Sci Educ 2014;7(1):12-8. doi:10.1002/ase.1361

124. Raiss P, Rettig O, Wolf S, Loew M, Kasten P. [Range of motion of shoulder and elbow in activities of daily life in 3D motion analysis]. Z Orthop Unfall 2007;145(4):493-8. doi:10.1055/s-2007-965468 [doi]

125. Ramsey ML. Distal biceps tendon injuries: diagnosis and management. J Am Acad Orthop Surg 1999;7(3):199-207. doi:10.5435/00124635-199905000-00006

126. Regan W, Lapner PC. Prospective evaluation of two diagnostic apprehension signs for posterolateral instability of the elbow. J Shoulder Elbow Surg 2006;15(3):344-6. doi:S1058-2746(05)00111-4 [pii];10.1016/j.jse.2005.03.009 [doi]

127. Reinold MM, Wilk KE, Macrina LC, Sheheane C, Dun S, Fleisig GS, et al. Changes in shoulder and elbow passive range of motion after pitching in professional baseball players. Am J Sports Med 2008;36(3):523-7. doi:10.1177/0363546507308935

128. Roebroek ME, Harlaar J, Lankhorst GJ. The application of generalization theory to reliability assesment: an illustration using isometric force measurements. Phys Ther 1993;73(6):386-95.

129. Rothstein JM, Miller PJ, Roettger RF. Goniometric reliability in a clinical setting. Elbow and knee measurements. Phys Ther 1983;63(10):1611-5.

130. Ruch DS, Papadonikolakis A, Campolattaro RM. The posterolateral plica: a cause of refractory lateral elbow pain. J Shoulder Elbow Surg 2006;15(3):367-70. doi:10.1016/j.jse.2005.08.013

131. Ruland RT, Dunbar RP, Bowen JD. The biceps squeeze test for diagnosis of distal biceps tendon ruptures. Clin Orthop Relat Res 2005(437):128-31. doi:00003086-200508000-00023 [pii]

132. Safran MR, McGarry MH, Shin S, Han S, Lee TQ. Effects of elbow flexion and forearm rotation on valgus laxity of the elbow. J Bone Joint Surg Am 2005;87(9):2065-74. doi:10.2106/JBJS.D.02045

133. Sardelli M, Tashjian RZ, MacWilliams BA. Functional elbow range of motion for contemporary tasks. J Bone Joint Surg Am 2011;93(5):471-7. doi:10.2106/JBJS.I.01633

134. Schenkels E, Caekebeke P, Swinnen L, Peeters J, van Riet R. Is the flexion-abduction-supination magnetic resonance imaging view more accurate than standard magnetic resonance imaging in detecting distal biceps pathology? J Shoulder Elbow Surg 2020;29(12):2654-60. doi:10.1016/j.jse.2020.05.014
135. Scholtes VA, Terwee CB, Poolman RW. What makes a measurement instrument valid and reliable? Injury 2011;42(3):236-40. doi:S0020-1383(10)00777-1 [pii];10.1016/j.injury.2010.11.042 [doi]
136. Seidenberg PH, Beutler A. The Sports Medicine Resource Manual. Elsevier; 2008.
137. Shaaban H, Pereira C, Williams R, Lees VC. The effect of elbow position on the range of supination and pronation of the forearm. J Hand Surg Eur Vol 2008;33(1):3-8. doi:33/1/3 [pii];10.1177/1753193407087862 [doi]
138. Shanley E, Rauh MJ, Michener LA, Ellenbecker TS, Garrison JC, Thigpen CA. Shoulder range of motion measures as risk factors for shoulder and elbow injuries in high school softball and baseball players. Am J Sports Med 2011;39(9):1997-2006. doi:10.1177/0363546511408876
139. Sharoff L. Integrating YouTube into the nursing curriculum. Online J Issues Nurs 2011;16(3):6. doi:10.3912/OJIN.Vol16No03PPT03
140. Shiba R, Sorbie C, Siu DW, Bryant JT, Cooke TD, Wevers HW. Geometry of the humeroulnar joint. J Orthop Res 1988;6(6):897-906. doi:10.1002/jor.1100060614 [doi]
141. Shim SS, Strauch RJ. A novel clinical test for partial tears of the distal biceps brachii tendon: The TILT sign. Clin Anat 2018;31(2):301-3. doi:10.1002/ca.23038
142. Simel DL, Samsa GP, Matchar DB. Likelihood ratios with confidence: sample size estimation for diagnostic test studies. J Clin Epidemiol 1991;44(8):763-70.
143. Simonsen O, Jensen J, Mouritsen P, Lauritzen J. The accuracy of clinical examination of injury of the knee joint. Injury 1984;16(2):96-101.
144. Skirven, Osterman, Fedorczyk, Amadio. Rehabilitation of the Hand and Upper Extremity. Elsevier; 2011.
145. Smith W, Rafeek R, Marchan S, Paryag A. The use of video-clips as a teaching aide. Eur J Dent Educ 2012;16(2):91-6. doi:10.1111/j.1600-0579.2011.00724.x
146. Solomon DH, Simel DL, Bates DW, Katz JN, Schaffer JL. The rational clinical examination. Does this patient have a torn meniscus or ligament of the knee? Value of the physical examination. JAMA 2001;286(13):1610-20.
147. Soucie JM, Wang C, Forsyth A, Funk S, Denny M, Roach KE, et al. Range of motion measurements: reference values and a database for comparison studies. Haemophilia 2011;17(3):500-7. doi:10.1111/j.1365-2516.2010.02399.x [doi]

148. Spinner M, Kaplan EB. The quadrate ligament of the elbow--its relationship to the stability of the proximal radio-ulnar joint. Acta Orthop Scand 1970;41(6):632-47.
149. Srivastava G, Roddy M, Langsam D, Agrawal D. An educational video improves technique in performance of pediatric lumbar punctures. Pediatr Emerg Care 2012;28(1):12-6. doi:10.1097/PEC.0b013e31823ed5ed
150. Stokdijk M, Biegstraaten M, Ormel W, de Boer YA, Veeger HE, Rozing PM. Determining the optimal flexion-extension axis of the elbow in vivo - a study of interobserver and intraobserver reliability. J Biomech 2000;33(9):1139-45. doi:S0021-9290(00)00079-8 [pii]
151. Urch E, Taylor SA, Cody E, Fabricant PD, Burket JC, O'Brien SJ, et al. The Quality of Open-Access Video-Based Orthopaedic Instructional Content for the Shoulder Physical Exam is Inconsistent. HSS J 2016;12(3):209-15. doi:10.1007/s11420-016-9508-6
152. Valdes K, LaStayo P. The value of provocative tests for the wrist and elbow: a literature review. J Hand Ther 2013;26(1):32-42; quiz 3. doi:10.1016/j.jht.2012.08.005
153. van de Pol RJ, van TE, Lucas C. Inter-rater reliability for measurement of passive physiological range of motion of upper extremity joints is better if instruments are used: a systematic review. J Physiother 2010;56(1):7-17.
154. van Rijn SF, Zwerus EL, Koenraadt KL, Jacobs WC, van den Bekerom MP, Eygendaal D. The reliability and validity of goniometric elbow measurements in adults: A systematic review of the literature. Shoulder Elbow 2018;10(4):274-84. doi:10.1177/1758573218774326
155. Werner BC, Holzgrefe RE, Griffin JW, Lyons ML, Cosgrove CT, Hart JM, et al. Validation of an innovative method of shoulder range-of-motion measurement using a smartphone clinometer application. J Shoulder Elbow Surg 2014;23(11):e275-82. doi:10.1016/j.jse.2014.02.030
156. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011;155(8):529-36. doi:155/8/529 [pii];10.7326/0003-4819-155-8-201110180-00009 [doi]
157. Wilk KE, Macrina LC, Fleisig GS, Aune KT, Porterfield RA, Harker P, et al. Deficits in glenohumeral passive range of motion increase risk of elbow injury in professional baseball pitchers: a prospective study. Am J Sports Med 2014;42(9):2075-81. doi:10.1177/0363546514538391
158. Wright RW, Steger-May K, Wasserlauf BL, O'Neal ME, Weinberg BW, Paletta GA. Elbow range of motion in professional baseball pitchers. Am J Sports Med 2006;34(2):190-3. doi:0363546505279921 [pii];10.1177/0363546505279921 [doi]
159. Zwerus EL, Somford MP, Maissan F, Heisen J, Eygendaal D, van den Bekerom MP. Physical examination of the elbow, what is the evidence? A systematic literature review. Br J Sports Med 2017. doi:10.1136/bjsports-2016-096712

160. Zwerus EL, Willigenburg NW, Scholtes VAB, Somford MP, Eygendaal D, van den Bekerom MPJ. Normative values and affecting factors for the elbow range of motion. Shoulder & Elbow 2017;[Epub ahead of print]. doi:10.1177/1758573217728711