Orthopaedics & Traumatology: Surgery & Research 105 (2019) 251–256

Contents lists available at ScienceDirect

Orthopaedics & Traumatology: Surgery & Research

journal homepage: www.elsevier.com

Original article

The "Auto-Constant": Can we estimate the Constant-Murley score with a self-administered questionnaire? A pilot study

Mikaël Chelli*, Yohann Levy, Vincent Lavoué, Gilles Clowez, Jean-François Gonzalez, Pascal Boileau

Institut Universitaire Locomoteur et du Sport, University Hospital of Nice, 06000 Nice, France

ARTICLE INFO

Article history:
Received 29 January 2018
Accepted 28 November 2018

Keywords:
Shoulder
Patient Reported Outcome Measures
Surveys and Questionnaires

ABSTRACT

Introduction: The Constant score, allows an objective and subjective assessment of the shoulder function. It has been proven to have a poor interobserver reliability for some of its aspects and is not usable as a remote assessment tool.
Hypothesis: The Constant-Murley functional shoulder score can be assessed with a self-administered questionnaire.
Methods: We conducted a prospective continuous study in a shoulder-specialized service. For each patient seen in consultation or hospitalized for a shoulder pathology, a self-administered questionnaire was delivered, and a clinical examination was performed by a surgeon. The questionnaire, in French language, was composed of checkboxes only, with pictures preferred over text for most items. Correlations with surgeon examination were assessed with the intraclass correlation coefficients, differences with the paired *t*-test.
Results: One hundred consecutive patients were analyzed. Correlation between the two scores was excellent (0.87), as were the range of motion and the pain subscores (0.85 and 0.78), good for the activity (0.69) and fair for the strength (0.57). The mean total score was 3 points lower for the self-administered questionnaire (CI95 [−5; −1]; $p < 0.01$). Activity and pain were not significantly different (−0.4/20 and −0.3/40; $p > 0.05$) but pain and force were slightly different (+0.8/15; −3.0/25; $p < 0.01$).
Conclusion: The Auto-Constant questionnaire in French is an excellent estimator of the Constant score, and of its pain and mobility sub-scores. It is less accurate for the evaluation of the strength, but differences between sub-scores compensate and allow its use in daily practice.
Level of proof: II, Prospective continuous clinical series.

1. Introduction

The Constant-Murley Score (CS) is a well-established functional score for the shoulder, described in 1987 [1], and widely used in international studies [2–6]. It is calculated after a clinical examination and is divided in 4 subscores : pain (15 points), activity (20 points), mobility (40 points) and strength (25 points) with a minimum of 0 and a maximum of 100 points. By opposition to other widely-used shoulder scores, like the American Shoulder and Elbow Surgeons Score (ASES) [7] or the Simple Shoulder Test (SST) [8], the use of the CS is limited and time-consuming by the lack of a self-administered questionnaire (SAQ). However, it allows a subjective and objective assessment, by combination of a clinical examination and a questionnaire, which is not possible with the ASES or the SST.

It has been shown that interobserver reliability was insufficient for as much as 12 of the 14 items of the CMS, and could lead to differences of 10 to 25 points between observers [9,10]. To avoid this inter-observer bias, we were interested by using a SAQ. Moreover, some patients may not or would not come for clinical follow-up, and we wanted a tool to assess these patients remotely.

Levy et al. have established a self-administered questionnaire in English [11], and proved it was a reliable tool for estimating the CS, but the assessment of strength requires the patient to use various objects of his choice at home. This evaluation of strength limits its use in daily practice in our eyes and makes it complicated for the patient.

The EMPRO (Evaluating Measures of Patient Reported Outcomes) [12] group evaluated patient-reported shoulder scores and found the ASES, the SST and the Oxford Shoulder Score to be reliable

* Corresponding author. University hospital of Nice, Institut Universitaire Locomoteur et du Sport, 30, voie Romaine, Nice 06000, France.
E-mail address: mikael@chel.li (M. Chelli).

https://doi.org/10.1016/j.otsr.2018.11.023
1877-0568/

and responsive. However, none of these scores can provide direct assessment of mobility or strength.

We wanted to combine the advantages of a self-administered questionnaire and those of a clinical-based functional score. Therefore, we created a self-administered questionnaire in French based upon the CS: The Auto-Constant questionnaire.

Our objective was to assess the correlation of such a questionnaire with the usual assessment by a clinical examination.

We hypothesized that a self-administered questionnaire was sufficiently efficient to estimate the Constant-Murley score, as compared to a clinical examination performed by a surgeon.

2. Methods

We studied a prospective continuous series of patients between January and February 2016 in our shoulder-specialized surgical unit. A standardized questionnaire in French language was established and given to all patients hospitalized or seen in clinics during the period of the study. Then, each patient was examined by a surgeon or a resident and a standard CS was calculated, blind to the results of the questionnaire.

2.1. Eligibility criteria

We included all the patients seen in clinics or hospitalized for a shoulder pathology, and for whom a self-administered questionnaire was available. We excluded patients with a missing questionnaire or having an incomplete clinical examination, patients below 18, or who did not speak French.

2.2. Questionnaire

The questionnaire was distributed before the clinical examination and did not require any assistance to be completed except for two patients having difficulty speaking French. All the questions could be answered with checkboxes without writing, to decrease the language bias and to ease the completion by the patient. We tried to be as close as possible to the original version proposed by Constant and Murley [1], and to the actualized recommendations by Constant et al. [13].

2.3. Pain (15 points)

A visual analogic scale (VAS) from 0 to 10 points was proposed to estimate the pain of the patient, asking the maximal level of pain felt on a common day [13]. Bounds of the VAS were entitled “No pain” for 0 and “Intolerable pain” for 10 (Supplementary Data page 1)[1]. Measures were rounded to the nearest integer value (3.4 would become 3).

2.4. Activity (20 points)

VAS from 0 to 10 points were also used for the questions concerning professional and leisure impairments to maximize the objectivity of answers, and as recommended by Constant et al. [13]. We accounted 0 point of impairment out of 4 if the VAS was below 1 (“No activity or No Work”, “No leisure”), 1 point for a VAS between 1 and 3, 2 points for a VAS between 4 and 6, 3 points for a VAS between 7 and 9, and 4 points for a VAS of 10 (“All my activities/leisures are possible”, Supplementary Data page 2).

For the question on the level of work with the hand, we used photographs to illustrate the different levels. During the study, we changed the title of the question because we felt empirically that patients were underestimating their level. Indeed, the initial question was “At what level can you work comfortably with your hand?” and patients were choosing the level at which they were the most comfortable which was usually the lowest. We simply used the plural version of the sentence, asking “At which level(s) . . .” instead, therefore requiring checking one or several cases. We then retained the highest answer if several answers were checked.

2.5. Mobility (40 points)

Photographs were used to allow auto-assessment by the patient (Supplementary Data pages 4–7). A healthy model was photographed in different positions of flexion, abduction and rotation according to the existing cutoffs of the CS. The question asked was “How can you move you shoulder WITHOUT pain?” according to the recommendations [13].

2.6. Strength (25 points)

We tried to estimate by a simple way the strength of patients, by asking what they felt they could lift and showing photographs of daily life objects, lifted by a healthy subject at 90 degrees of abduction (Supplementary Data page 3). Initially, the question was “Which weight do you think you can carry at arm’s length?” to be the closest to the assessment of strength in the CS, which must be done at 90° of abduction. However, we quickly noted that patients were heavily underestimating their strength, because most of them thought they could not carry anything at arm’s length, at 90° of abduction due to their pain, which was wrong when examining them. We therefore changed the question to “Which weight do you think you can lift?” and removed the absolute weight in kilograms below the photographs but kept the pictures at 90° of abduction to stay loyal to the original score.

3. Clinical Examination

After the SAQ was completed, each patient was examined by a shoulder-specialized orthopedic surgery fellow. No training or specific instructions were given to the participants (patients or physician), in order to stay in the conditions of our everyday practice, and examiners were blind to the result of the SAQ. A mechanical spring balance was used in all cases to measure strength in 90° of abduction and with the arm in the axis of the scapula. Data were recorded electronically in our patient management software.

3.1. Flow chart

One hundred and seventy-one patients completed a questionnaire during the time of the study. The main reason for exclusion was a missing clinical examination by a surgeon in 59 cases (Fig. 1). We finally included the 100 first patients in a continuous way, and then stopped the study.

3.2. Statistical analysis

Absolute differences between surgeon’s score and self-administered questionnaire were tested with the paired samples t-test. Positive differences were related to an overestimation by the patient and negative differences expressed an underestimation by the patient as compared to the surgeon’s examination. Alpha risk was established as $p = 0.05$.

The correlation between the two scores was assessed with the intraclass correlation coefficient and Spearman’s rank coefficient as

[1] The online form, available in the electronic appendix, is also on: https://www.chu-nice.fr/institut-universitaire-locomoteur/orthopedie-chirurgie-sport/auto-constant/

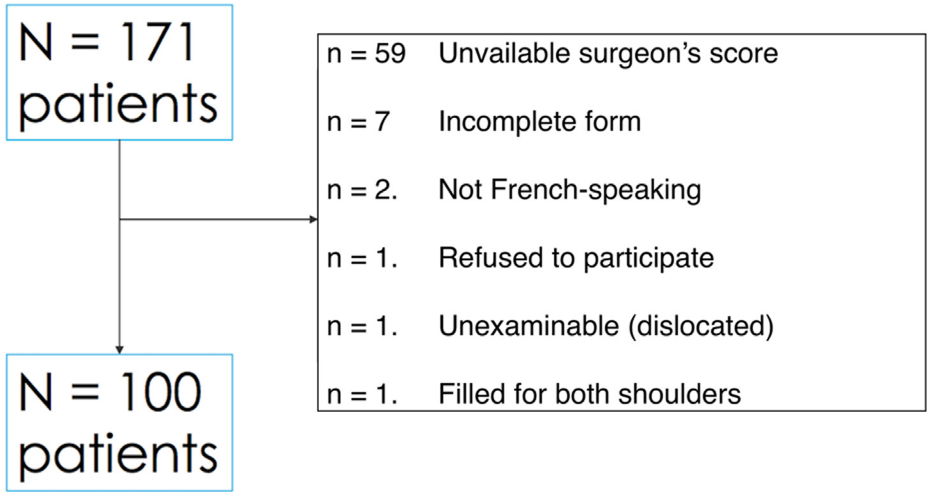

Fig. 1. Flow chart.

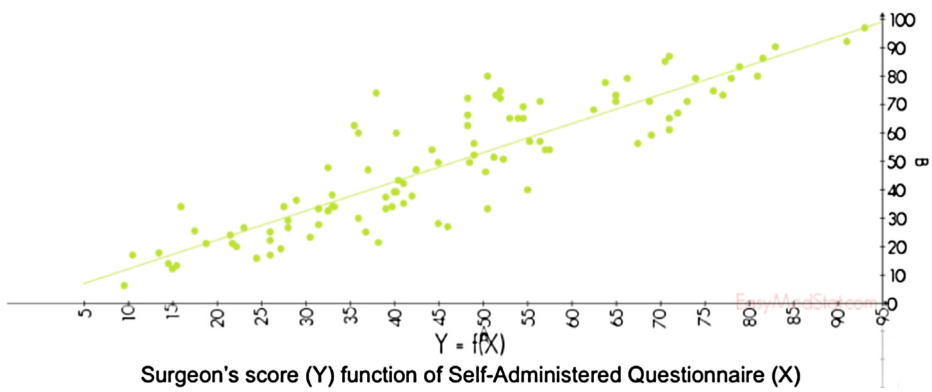

Fig. 2. Scatter plot of CS as calculated by the surgeon (ordinates) function of the score estimated by the self-administered questionnaire (abscesses).

some variables were not normally distributed (Shapiro-Wilk test). Intraclass correlations r were judged weak if between 0 and 0.39; passable between 0.40 and 0.59; good between 0.60 and 0.74; and excellent between 0.75 and 1, according to the recommendations of Ciccheti [14].

Data collection and statistical analysis were performed with the online application EasyMedStat (www.easymedstat.com; Neuilly-Sur-Seine; France).

3.3. Patients

One hundred patients (50 women and 50 men) were examined at a mean age of 57 years (Min. 18, Max. 90, SD 20.2). Thirty-eight had a rotator cuff condition (cuff tear or cuff tendinopathy), 24 had a shoulder osteoarthritis treated by arthroplasty, 16 suffered gleno-humeral anterior instability (due to glenoid bone loss or labral lesion), 1 patient had a non-operated primary osteoarthritis and 1 had a retractile capsulitis. Ninety patients were right-handed and 61 were consulting for the dominant side. The mean CS assessed by the surgeon was 49 with scores varying from 6 to 97.

4. Results

4.1. Correlations

The correlation between the two scores was excellent (r = 0.87) (Fig. 2). The correlation was excellent for the sub-scores Mobility and Pain (r = 0.85 and 0.78), good for the Activity (r = 0.69) and passable for the Strength (r = 0.57) (Table 1).

Table 1
Correlations between questionnaire and surgeon's scores.

Score	Intraclass correlation (r)	Spearman (r)
Total Score/100	0.87	0.88
Pain/15	0.78	0.79
Activity/20	0.69	0.70
Mobility/40	0.85	0.85
Strength/25	0.57	0.80

4.2. Absolute differences

The score was averagely underestimated of -3 points by the patients when answering to the SAQ ($p < 0.01$) with a 95% confidence interval between −5.1 and −1 points (Table 2). Activity and Mobility were not significantly different between the self-administered questionnaire and the surgeon's examination (−0.4/20 and −0.3/40; $p > 0.05$), but Pain and Strength were significantly different (+0.8/15 and −3.0/25; $p < 0.01$).

4.3. Influence of pathology

The correlation was excellent for the 3 main pathologies: 0.87 for osteoarthritis, 0.81 for rotator cuff pathology and 0.80 for anterior instability (Table 3).

Table 2
Mean differences between questionnaire and surgeon's scores (negative values represent underestimation of the score by the patient as compared to the surgeon's examination).

Parameter	Mean difference	95% Confidence Interval	p-Value
Total Score/100	−3.0	−5.1; −1	< 0.01
Pain/15	+0.8	+ 1.3; + 0.3	< 0.01
Activity/20	−0.4	− 1.1; + 0.2	NS
Mobility/40	−0.3	− 1.5; + 0.8	NS
Strength/25	−3.0	− 4.0; − 2.0	< 0.01

NS: not significant.

Table 3
Influence of pathology.

Pathology	Intraclass correlation (r)	Spearman (r)
Cuff lesion $n = 38$	0.81	0.82
Arthroplasty $n = 24$	0.87	0.85
Instability $n = 16$	0.80	0.87

Table 4
Influence of examiner.

Examiner	Intraclass correlation (r)	Spearman (r)
MC $n = 52$	0.93	0.93
Other $n = 48$	0.79	0.80

4.4. Influence of examiner

Fifty-two patients were examined by the main author (MC) with a correlation r = 0.93, while 48 patients were examined by other examiners with a correlation r = 0.79, both excellent (Table 4).

4.5. Influence of formulation

4.5.1. Hand level

For the first formulation (singular, $n = 54$), the correlation was passable (r = 0.44) and the mean difference of 1.64/10. For the second formulation (plural, $n = 46$), the correlation was good (r = 0.64) and the mean difference of 0.67/10 (Table 5).

4.5.2. Strength

In the first version of the questionnaire ($n = 32$), the correlation was passable (r = 0.47) and the mean difference of 4.1/25. For the second version ($n = 68$), the correlation was good (r = 0.67) with a mean difference of 2.5/25 (Table 6).

4.6. Floor or ceiling effect

No patient reached the minimum (0) or maximum value (100), either after the surgical examination or with the self-administered questionnaire. Therefore, we did not observe floor or ceiling effect.

5. Discussion

Our results confirm our hypothesis that a self-administered questionnaire can estimate precisely the CS. The 95% confidence interval of the difference is between −1 and −5 points, with the patients underestimating their score. Moreover, the consistency of the correlations among the different pathologies lets us think that this questionnaire could be used in most of common shoulder pathologies, although instability should be assessed with complementary specific tools [15].

The compliance of the patients to complete this questionnaire was high since 163 out of 171 patients have correctly completed it (95%). However, among these 163 patients, 59 did not have a complete clinical examination by a surgeon. This lets us think that a self-administered questionnaire could improve our exhaustivity in data collection [16].

We could not find an excellent correlation for two of the subscores of the CS. Only pain and mobility items could be assessed precisely by the questionnaire. Eventually, differences between sub-scores compensate with a slight overestimation of the pain by the questionnaire, leading to an excellent global correlation between the total scores.

A systematic review of literature in 2010 [15] has shown that the inter- and intra-observer reliability of the total CS was good for several pathologies [9,10], but this reliability decreases when regarding subscores of the CS [17–22], especially for pain and impairment in daily life, which are the two subjective items of this score. But as we observed with our questionnaire, these differences between subscores compensate when looking at the total score.

Rüdiger et al. [23] conducted a similar study with a SAQ in order to assess remotely shoulder range of motion. They asked patients to draw lines representing their mobility on prepared diagrams showing shoulders. Authors found a weak correlation between surgeon and patient assessment of mobility, with patients overestimating their range of motion. In our study we could not find a significant difference between the mobility subscore assessed by the patient or the surgeon. Using photographs instead of diagrams may have been easier for patients and led to a higher correlation.

Our results are below those reported by the self-administered questionnaire in English proposed by Levy et al. [11] regarding the global score, since their mean difference was -1.3 points (vs −3 points for our questionnaire) but with a wider confidence interval (−7 to +4 points), while our confidence interval is narrower and only contains negative values (−5 to −1 points). This means that with our questionnaire, in 95% of cases, the difference would be below 5 points, and always in the same direction (underestimation by the patient).

The difference between the main examiner (MC; r = 0.93) and the other examiners (r = 0.79) is probably multifactorial: an author−examiner bias of measurement, and the lack of standardization could be involved. However, we tried to limit this examiner bias by doing the clinical examination after the questionnaire and blind to the results of this questionnaire. For the lack of standardization, we voluntarily did not realize a specific formation of examiners for this study, to stay in the conditions of daily practice. As the main examiner was aware of the exact version of the questionnaire when designing the questionnaire, we expected a higher correlation for him, as found in our study. We could therefore expect a higher correlation, like the main examiner, if all examiners were trained.

The assessment of strength was the most biased in the SAQ, as we had predicted when designing it. Indeed, to propose a simple questionnaire we chose to approximate the evaluation of strength with a subjective question. We proposed only 5 objects of daily life to assess the strength, where the original evaluation is between 0 and 25 pounds, leading to 26 different values against 6 in our questionnaire (5 objects and "I cannot hold anything"). This part of the questionnaire should be improved by allowing more options and intermediate weights, to decrease the mean difference of −3 points for this question. The maximum weight proposed was a 6−pack of 1.5-liter water bottles which is roughly 18 pounds. In our usual practice, we rarely have patients able to lift more than 18 pounds at 90 degrees of abduction, but a higher maximum weight could be required for some patients.

Surgeons measured a strength below 10 pounds in 70% of patients. We had anticipated this and chose a larger number of light objects, rather than heavy objects. We could probably improve our questionnaire by proposing more light objects and asking the patients to choose all the objects he feels he can lift. We could even

Table 5
Influence of formulation on the question "Level of work with hand".

Version	Intraclass correlation (r)	Spearman	Mean difference
"At what level" (« A quel niveau ») n = 54	0.44	0.49	1.64/10
"At which level(s)"("A quels niveaux")n = 48	0.64	0.61	0.67/10

Table 6
Influence of formulation on the evaluation of strength.

Version	Intraclass correlation (r)	Spearman	Mean difference
"Carry at arm's length" + weights of objects in kg ("Porter à bout de bras")n = 32	0.47	0.87	4.1/25
"Lift" without weights ("Soulever") n = 68	0.64	0.78	2.5/25

add several objects of the same weight because the evaluation of the weight of an object by the patient could be biased.

Finally, it has been shown that this evaluation of strength is inherently subject to an inter-observer variability, even with a standardized measure [9,10]. This may be another factor affecting negatively the correlation with the SAQ.

Limitations of this study include the assessment of strength which should be improved and the changing titles of questions Hand Level and Strength. Indeed, these changes were not planned prior to the study, and could have introduced a bias in our results. However, it would be interesting to randomly test different versions of some of the questions to improve the reliability of each question independently, in a future version. Randomization would then limit the possible selection bias. Finally, examiners were not trained, and the clinical examination was not standardized. This limits the reproducibility of our results but was done on purpose for this pilot study.

Some parameters still need to be verified in order to completely assess this self-administered questionnaire. A third evaluation by a second examiner could be interesting to compare the inter-observer correlation with the questionnaire-surgeon correlation. This would allow to assess its external reliability. Responsiveness–the ability of accurately detect change over time–could be assessed by performing preoperative and postoperative scores in the same patients [15,16,24]. This questionnaire has been established and studied in French. An English version is currently being produced. Eventually, it should be compared to other patient-reported outcomes scores with high responsiveness [15] like the ASES in rotator cuff and arthritis pathologies or the Rowe score for instability.

6. Conclusion

The Auto-Constant questionnaire in French is an excellent global estimator of the Constant-Murley score, as well as the pain and mobility sub-scores, for the rotator cuff pathologies, shoulder osteoarthritis and shoulder instability. It still has some limitations for the evaluation of strength but differences between sub-scores compensate to finally allow a satisfying global correlation. Some modifications could allow this tool to be even more precise and are currently studied.

Disclosure of interest

The authors declare that they have no competing interest.

Funding

None.

Contribution

Mikaël Chelli designed the questionnaire. Pascal Boileau and Jean-François Gonzalez mentored the study, and helped writing the article. Other authors were the examiners.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.otsr.2018.11.023.

References

[1] Constant CR, Murley AH. A clinical method of functional assessment of the shoulder. Clin Orthop 1987;214:160–4.
[2] Bigorre N, Lancigu R, Bizot P, Hubert L. Predictive factors of scapular notching in patients with reverse shoulder arthroplasty. Orthop Traumatol Surg Res 2014;100:711–4.
[3] Gadea F, et al. Fixation of 4-part fractures of the proximal humerus: Can we identify radiological criteria that support locking plates or IM nailing?Comparative, retrospective study of 107 cases. Orthop Traumatol Surg Res 2016;102:963–70.
[4] Gallusser N, Farron A. Complications of shoulder arthroplasty for osteoarthritis with posterior glenoid wear. Orthop Traumatol Surg Res 2014;100:503–8.
[5] Boileau P, et al. Isolated loss of active external rotation: a distinct entity and results of L'Episcopo tendon transfer. J Shoulder Elbow Surg 2017, http://dx.doi.org/10.1016/j.jse.2017.07.008.
[6] Ascione F, et al. Long-term humeral complications after Grammont-style reverse shoulder arthroplasty. J Shoulder Elbow Surg 2018, http://dx.doi.org/10.1016/j.jse.2017.11.028.
[7] Michener LA, McClure PW, Sennett BJ. American Shoulder and Elbow Surgeons Standardized Shoulder Assessment Form, patient self-report section: reliability, validity, and responsiveness. J Shoulder Elbow Surg 2002;11:587–94.
[8] Lippitt SB, et al. A practical tool for evaluating function: The simple shoulder test; 1993.
[9] Roy J-S, MacDermid JC, Woodhouse LJ. A systematic review of the psychometric properties of the Constant-Murley score. J Shoulder Elbow Surg 2010;19:157–64.
[10] Rocourt MHH, et al. Evaluation of intratester and intertester reliability of the Constant-Murley shoulder assessment. J Shoulder Elbow Surg 2008;17:364–9.
[11] Levy O, Haddo O, Massoud S, Mullett H, Atoun E. A Patient-derived Constant-Murley Score is Comparable to a Clinician-derived Score. Clin Orthop 2014;472:294–303.
[12] Schmidt S, et al. Evaluation of shoulder-specific patient-reported outcome measures: a systematic and standardized comparison of available evidence. J Shoulder Elbow Surg 2014;23:434–44.
[13] Constant CR, et al. A review of the Constant score: Modifications and guidelines for its use. J Shoulder Elbow Surg 2008;17:355–61.
[14] Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychol Assess 1994;6:284–90.
[15] Unger RZ, et al. The Responsiveness of Patient- Reported Outcome Tools in Shoulder Surgery Is Dependent on the Underlying Pathological Condition. Am J Sports Med 2018, http://dx.doi.org/10.1177/0363546517749213 [036354651774921].
[16] Schamber EM, Takemoto SK, Chenok KE, Bozic KJ. Barriers to completion of Patient Reported Outcome Measures. J Arthroplasty 2013;28:1449–53.
[17] Ban I, Troelsen A, Christiansen DH, Svendsen SW, Kristensen MT. Standardised test protocol (Constant Score) for evaluation of functionality in patients with shoulder disorders. Dan Med J 2013;60:A4608.
[18] Livain T, et al. Intra- and interobserver reproducibility of the French version of the Constant-Murley shoulder assessment during rehabilitation after rotator cuff surgery. Rev Chir Orthop Reparatrice Appar Mot 2007;93:142–9.

[19] Blonna D, et al. Can we improve the reliability of the Constant-Murley score? J Shoulder Elbow Surg 2012;21:4–12.
[20] Hirschmann MT, Wind B, Amsler F, Gross T. Reliability of Shoulder Abduction Strength Measure for the Constant-Murley Score. Clin Orthop 2010;468:1565–71.
[21] Walton MJ, Walton JC, Honorez LAM, Harding VF, Wallace WA. A comparison of methods for shoulder strength assessment and analysis of Constant score change in patients aged over fifty years in the United Kingdom. J Shoulder Elbow Surg 2007;16:285–9.
[22] Johansson KM, Adolfsson LE. Intraobserver and interobserver reliability for the strength test in the Constant-Murley shoulder assessment. J Shoulder Elbow Surg 2005;14:273–8.
[23] Rüdiger HA, Fuchs B, von Campe A, Gerber C. Measurements of shoulder mobility by patient and surgeon correlate poorly: A prospective study. J Shoulder Elbow Surg 2008;17:255–60.
[24] Hawkins RJ. Recommendations for evaluating and selecting appropriately valued outcome measures. Instr Course Lect 2016;65:587–91.