Validity, responsiveness and minimal important change of the EQ-5D-5L in patients after rotator cuff repair, shoulder arthroplasty or thumb carpometacarpal arthroplasty

Miriam Marks[1] · Cécile Grobet[1] · Laurent Audigé[1]

Abstract

Purpose The aim was to investigate the measurement properties of the EQ-5D-5L utility index in patients after arthroscopic rotator cuff repair (RCR), total shoulder arthroplasty (TSA) or thumb carpometacarpal (CMC I) arthroplasty.

Methods In this prospective study, all patients completed the EQ-5D-5L before surgery and 6 months and 1 year after surgery. In addition, RCR patients completed the Oxford Shoulder Score (OSS), TSA patients completed the Shoulder Pain and Disability Index (SPADI) and CMC I patients completed the brief Michigan Hand Outcomes Questionnaire (brief MHQ) at each designated time point. Construct validity (Pearson's correlation coefficient, r), responsiveness (effect size), minimal important difference (MID), minimal important change (MIC), and floor and ceiling effects of the EQ-5D-5L were determined. To test discriminative ability, EQ-5D-5L utility indices of patients who were in a patient acceptable symptom state (PASS) or not at follow-up were compared using the Mann–Whitney U test.

Results We included 153 RCR, 150 TSA, and 151 CMC I patients. The EQ-5D-5L utility index correlated with the OSS ($r = 0.73$), SPADI ($r = -0.65$) and brief MHQ ($r = 0.61$). The effect sizes were 1.3 (RCR and CMC I group) and 1.1 (TSA). The MID and MIC ranged from 0.027 to 0.209. Ceiling effects were found. The EQ-5D-5L utility index differed significantly between patients being in a PASS versus patients who were not in a PASS.

Conclusion The EQ-5D-5L utility index shows good construct validity, responsiveness and discriminative ability in patients after arthroscopic RCR, TSA and CMC I arthroplasty and is suitable to quantify quality of life.

Clinical trial registration: This auxiliary analysis is part of a primary study that was originally registered at ClinicalTrials.gov (NCT01954433) on October 1, 2013.

Plain English summary

The EQ-5D-5L questionnaire is a common tool used to measure health-related quality of life. It is one of the shortest and simplest instruments used to monitor health. However, its performance (measurement properties) in upper-extremity orthopaedic conditions has not been fully evaluated, yet. In our analysis, we studied the performance of the ED-5D-5L in patients who had surgery for various shoulder or hand conditions. Patients completed the EQ-5D-5L and other questionnaires documenting functional scores before surgery, 6 months and 1 year after surgery. We found that the EQ-5D-5L has good measurement characteristics and is suitable for assessing general quality of life in our patients with shoulder or hand disorders.

Introduction

At the upper extremity, there are three prevalent musculoskeletal disorders that can markedly affect patient quality of life (QoL): rotator cuff tears as a result of direct trauma or age-related degenerative changes, glenohumeral osteoarthritis and thumb carpometacarpal (CMC I) osteoarthritis. Surgical treatment for these three specific conditions includes rotator cuff repair (RCR), total shoulder

✉ Miriam Marks
 miriam.marks@kws.ch

1 Department of Teaching, Research and Development, Schulthess Klinik, Lengghalde 2, 8008 Zurich, Switzerland

arthroplasty (TSA) and CMC I arthroplasty, respectively, all of which lead to good functional restoration of the joint and increased QoL for affected patients [1–3].

Since economic evaluations of interventions for musculoskeletal disorders are increasingly important nowadays, a sound outcome measure for quantifying QoL and quality-adjusted life years (QALYs) is required. Quality of life assessments for the aforementioned patient groups are also increasing, yet there is a lack of knowledge regarding the performance of QoL measures such as the EQ-5D and just how adequate this tool, in particular, is in determining health-related QoL.

For musculoskeletal disorders of the upper extremity, there are only a few studies that have investigated the measurement properties of the EQ-5D. For patients with proximal humeral fractures and carpal tunnel syndrome, the EQ-5D shows good reliability and construct validity [4–7]. Responsiveness was found to be high in patients after various elective shoulder surgery procedures [8, 9], but only moderate to good after interventions targeting the hand [6, 10, 11].

The aim of our study was to investigate the measurement properties of the EQ-5D-5L utility index in patients after arthroscopic RCR, TSA and CMC I arthroplasty. In particular, we focused on evaluating the properties of construct validity, responsiveness, minimal important difference (MID), minimal important change (MIC), floor and ceiling effects and discriminative ability.

Methods

This analysis forms part of a previously published prospective cost-utility study [12, 13], which was conducted in a Swiss orthopaedic hospital and approved by the local ethics committee. We adhered to the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement [14].
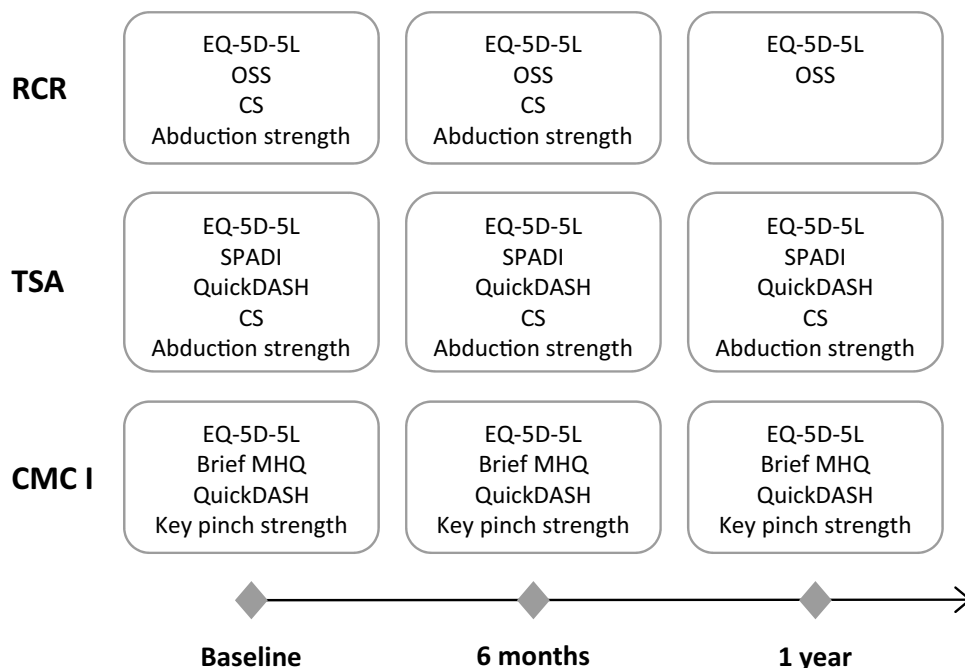
Patients

The study included three cohorts of: 1. patients with a rotator cuff tear who underwent arthroscopic RCR, 2. patients with glenohumeral osteoarthritis and/or rotator cuff tear who received a TSA, and 3. patients with CMC I osteoarthritis who underwent CMC I arthroplasty. All patients gave written informed consent for participation.

Assessments

Patients completed a set of German language questionnaires and underwent a clinical examination before surgery (i.e. baseline), and 6 months and 1 year post-surgery. RCR patients were not required to attend a clinical examination at 1 year. The timing of the following outcome measures is shown in Fig. 1.

After completion of the EQ-5D-5L [15], the 5 responses of the descriptive system were converted into an utility index using the EQ-5D-5L value set for Germany, which ranges from − 0.66 (lowest) to 1 (highest QoL) [16].

Fig. 1 Outcome measures for the three patient groups at the different time points. *RCR* rotator cuff repair, *TSA* total shoulder arthroplasty, *CMC I* thumb carpometacarpal arthroplasty, *OSS* Oxford Shoulder Score, *CS* Constant Score, *SPADI* Shoulder Pain and Disability Index, *QuickDASH* short version of the Disabilities of Arm, Shoulder and Hand Questionnaire, *Brief MHQ* brief Michigan Hand Outcomes Questionnaire

RCR patients completed the condition-specific Oxford Shoulder Score (OSS) [17], which ranges from 0 (worst) to 48 (best outcome) based on subjective levels of pain and impairment during activities of daily living.

Shoulder function of TSA patients was assessed using the Shoulder Pain and Disability Index (SPADI) [18], where scores range from 0 to 100 with higher scores indicating greater disability.

CMC I patients completed the brief Michigan Hand Outcomes Questionnaire (brief MHQ) for measuring hand function [19]. The score ranges from 0 to 100 with higher scores indicating better hand function.

TSA and CMC I patients also reported their function of the upper extremity by completing the short version of the Disabilities of the Arm, Shoulder and Hand questionnaire (QuickDASH) [20]. The QuickDASH yields a summary score between 0 and 100 with 0 indicating the best score or the lowest level of disability.

At both follow-ups, all patients were asked about their perceived change regarding their QoL, as has been proposed in the literature [21–23]: "In reference to your quality of life, do you feel *much better/slightly better/unchanged/slightly worse/much worse* compared to before the surgery?"

The two shoulder cohorts underwent clinical examinations for muscle strength in 90° abduction measured with a spring balance (Pesola AG, Schindellegi, Switzerland) and overall shoulder function based on the Constant Score (CS) [24]. The latter objectively measures range of motion, strength, pain and use of the shoulder in performing daily activities to yield a summary score between 0 (worst) and 100 (best outcome).

The clinical examination for CMC I patients included key pinch strength as measured in a standardized sitting position using a pinch gauge (B&L Engineering, Santa Ana, California, United States of America).

Statistics

Standard descriptive statistics with means and standard deviations (SDs) were presented. To test for changes in outcomes between baseline and the 6-month and 1-year follow-ups, we used ANOVA for repeated measures. For assessing the clinical outcomes of the RCR group only recorded at 6 months, the Wilcoxon signed-rank test was used. As there were only a few missing data, they were not replaced.

The evaluation of measurement properties was based on the definitions and recommendations of the COnsensus-based Standards for the selection of health status Measurement INstruments (COSMIN) group [21, 25, 26].

Construct validity was tested with the following predefined hypotheses using the Pearson's correlation coefficient (r):

a. There is at least a moderate correlation of larger than 0.6 between the EQ-5D-5L utility index and OSS in the RCR group.

b. There is at least a moderate negative correlation of larger than -0.6 between the EQ-5D-5L utility index and SPADI in the TSA group.

c. There is at least a moderate correlation of larger than 0.6 between the EQ-5D-5L utility index and brief MHQ score in the CMC I group.

Responsiveness was evaluated by testing the predefined hypothesis that the effect size (Cohen's d [27]) of the EQ-5D-5L utility index is greater than or equal to 0.8 in all three cohorts. This measurement property was calculated for the 1-year follow-up.

For *interpretability,* we calculated the minimal important difference (MID), which evaluates the smallest differences between patients or groups that are considered important [28]. The MID was determined by the difference in the EQ-5D utility index of patients who indicated a *slightly better* QoL after surgery and those with an *unchanged* QoL. Furthermore, the minimal important change (MIC), which defines the smallest change that patients consider important [28], was calculated using the following three methods: 1. a distribution-based method in which the MIC is equivalent to 0.5*SD [21, 29, 30], 2. an anchor-based approach using the mean change method and 3. another anchor-based approach using receiver operating characteristics (ROC). For the second method, the question about perceived change in QoL at follow-up compared with the baseline status was used as the anchor; the MIC was reflected by the mean change in the EQ-5D-5L utility index for those patients who considered themselves as feeling slightly better than before the surgery. For the ROC method, the 5-point Likert scale anchor question was transformed to a dichotomous scale with patients who had answered *much better* or *slightly better* being allocated to the improved group, and those with *unchanged*, *slightly worse* or much *worse* answers allocated to the comparison group of unimproved participants. The MIC was determined by finding the optimal cut-off point on the ROC curves, namely the point at which (1-sensitivity)+(1-specificity) was smallest [21].

Floor and ceiling effects at 1 year were noted if more than 15% of the patients achieved the lowest or highest possible scores [31].

To test the *discriminative ability* of the EQ-5D-5L, patients were divided into two groups based on whether they were either in a patient acceptable symptom state (PASS) at follow-up or not. The PASS is a patient's subjective rating on whether they feel good and are satisfied with their current health status or not. Statistically, it is the value in a score beyond which patients consider themselves well [32]. For the RCR group, patients were

categorized based on the OSS score. Christie et al. [33] published the PASS value of 26 for the OSS based on the former scoring system, which now corresponds to a value of 34 using the current OSS scoring method that we applied in this analysis [17]. The TSA patients were distributed based on the PASS value of 33.7 for SPADI [33], and CMC I patients based on the PASS value of 70 points for the brief MHQ [34]. The Mann–Whitney U test was used to test for differences in the EQ-5D-5L utility index for patients in either a PASS or not.

Results

From a total of 454 included patients, there were 153 in the RCR group, 150 in the TSA group and 151 in the CMC I group (Table 1). At the 1-year follow-up, the EQ-5D-5L utility indices of 148 RCR, 148 TSA and 150 CMC I patients were available. At baseline, the most affected EQ-5D-5L dimension for all patients was pain. Eighty percent of the patients indicated that they had moderate pain or worse (Fig. 2). For all patients, the mean baseline EQ-5D-5L utility index increased from 0.69 (SD 0.22) to 0.89 (SD 0.15) at 6 months and 0.90 (SD 0.13) at 1 year

Table 1 Patient characteristics

	All	RCR	TSA	CMC I
Patients; n	454	153	150	151
Age at surgery; mean (SD)	64 (10)	57 (8)	71 (9)	65 (8)
Gender, female; n (%)	259 (57)	56 (36)	86 (57)	117 (77)
Type of surgery; n (%)*				
Arthroscopic rotator cuff repair		153 (100)		
Anatomical TSA			47 (31)	
Reverse TSA			103 (69)	
Simple trapeziectomy				10 (7)
Trapeziectomy with LRTI (autograft)				127 (84)
Trapeziectomy with LRTI (allograft)				7 (5)
CMC I pyrocardan implant				6 (4)
Trapeziectomy with absorbable gelatine sponge				1 (1)

RCR rotator cuff repair, *TSA* total shoulder arthroplasty, *CMC I* thumb carpometacarpal joint arthroplasty, *LRTI* ligament reconstruction and tendon interposition

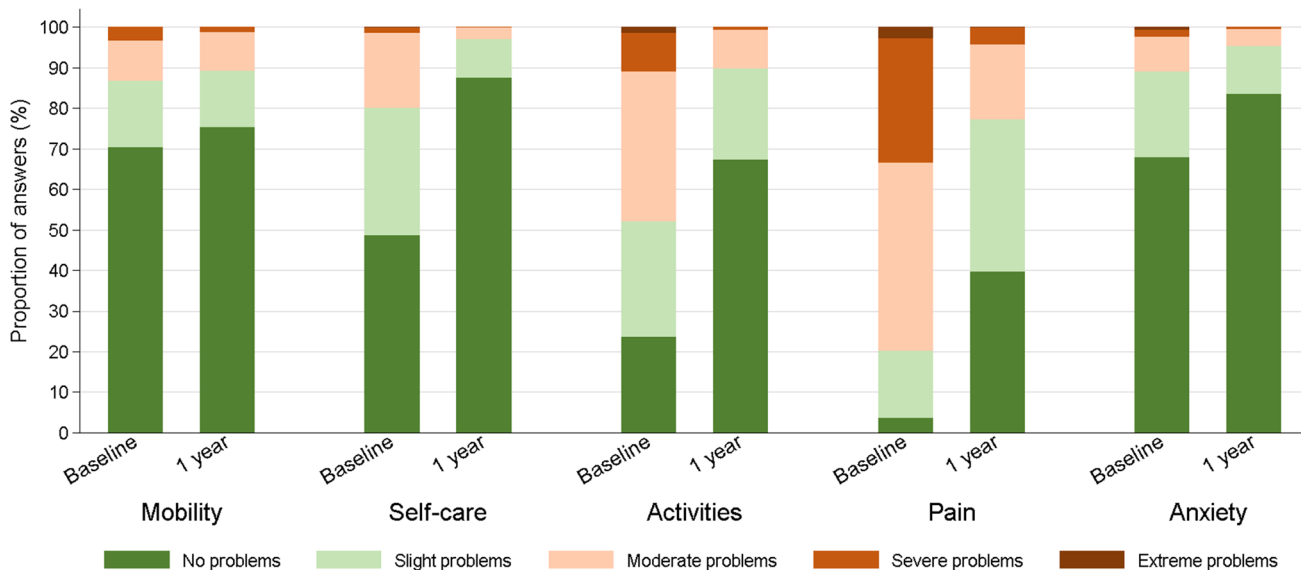*Percentages may differ from 100 due to rounding errors

Fig. 2 Dimension-specific results of the EQ-5D-5L for all patients at baseline and 1-year follow-up

($p \leq 0.001$). There was also a significant increase in all outcome measures for each group, except key pinch strength (Table 2).

The EQ-5D-5L utility index showed moderate to high correlations with the OSS ($r=0.73$), SPADI ($r=-0.65$) and brief MHQ ($r=0.61$) (Tables 3, 4 and 5).

Table 2 Outcome measures at baseline, 6 months and 1 year stratified by intervention group. Means and standard deviations are shown

	RCR	TSA	CMC I
EQ-5D-5L utility index [−0.66 to 1][a]			
Baseline	0.71 (0.23)	0.68 (0.23)	0.69 (0.19)
6 months	0.91 (0.14)	0.88 (0.15)	0.87 (0.15)
1 year	0.94 (0.11)	0.89 (0.15)	0.88 (0.11)
p-value	≤0.001	≤0.001	≤0.001
Oxford Shoulder Score [0 to 48][a]			
Baseline	27 (8)		
6 months	42 (7)		
1 year	44 (5)		
p-value	≤0.001		
SPADI [0 to 100][b]			
Baseline		62 (20)	
6 months		18 (18)	
1 year		15 (18)	
p-value		≤0.001	
Brief MHQ [0 to 100][a]			
Baseline			49 (15)
6 months			77 (19)
1 year			83 (16)
p-value			≤0.001
QuickDASH [0 to 100][b]			
Baseline		48 (17)	47 (16)
6 months		22 (18)	24 (19)
1 year		20 (17)	20 (17)
p-value		≤0.001	≤0.001
Constant Score [0 to 100][a]			
Baseline	48 (17)	35 (15)	
6 months	69 (14)	68 (13)	
1 year		71 (15)	
p-value	≤0.001	≤0.001	
Abduction strength [kg]			
Baseline	4.0 (4.0)	1.0 (3.0)	
6 months	7.0 (3.0)	5.0 (3.0)	
1 year		5.6 (2.9)	
p-value	≤0.001	≤0.001	
Key pinch strength [kg]			
Baseline			3.9 (2.0)
6 months			3.4 (1.5)
1 year			3.9 (1.7)
p-value			0.15

p-values report the results of ANOVA for repeated measures (two follow-ups) or the Wilcoxon signed-rank test (one follow-up)

RCR rotator cuff repair, *TSA* total shoulder arthroplasty, *CMC I* thumb carpometacarpal joint arthroplasty, *SPADI* Shoulder Pain and Disability Index, *Brief MHQ* brief Michigan Hand Outcomes Questionnaire, *QuickDASH* short version of the Disabilities of the Arm, Shoulder and Hand questionnaire

[a] Higher values indicate better function

[b] Higher values indicate worse function

Table 3 Pearson's correlation coefficients for the RCR group

	EQ-5D-5L utility index	OSS	CS
Oxford Shoulder Score (OSS)	0.73		
Constant Score (CS)	0.64	0.81	
Abduction strength	0.38	0.53	0.82

All correlations are significant with $p < 0.001$

Table 4 Pearson's correlation coefficients for the TSA group

	EQ-5D-5L utility index	QuickDASH	SPADI	CS
QuickDASH	− 0.67			
SPADI	− 0.65	0.86		
Constant Score (CS)	0.61	− 0.80	− 0.90	
Abduction strength	0.41	− 0.61	− 0.65	0.83

All correlations are significant with $p < 0.001$

QuickDASH short version of the Disabilities of the Arm, Shoulder and Hand questionnaire, *SPADI* Shoulder Pain and Disability Index

Table 5 Pearson's correlation coefficients for the CMC I group

	EQ-5D-5L utility index	QuickDASH	Brief MHQ
QuickDASH	− 0.64		
Brief MHQ	0.61	− 0.87	
Pinch grip	0.22	− 0.34	0.30

All correlations are significant with $p < 0.001$

QuickDASH short version of the Disabilities of the Arm, Shoulder and Hand questionnaire, *brief MHQ* brief Michigan Hand Outcomes Questionnaire

Responsiveness of the EQ-5D-5L was high in all three groups with effect sizes of 1.3 (RCR and CMC I group) and 1.1 (TSA) (Table 6). All condition-specific patient-reported outcomes showed higher effect sizes than the EQ-5D-5L.

The MID and MIC values, which ranged from 0.027 to 0.209, varied depending on the intervention group and also the calculation method for the latter (Table 7). With the ROC method, the area under the curve (AUC) values were 0.75, 0.69 and 0.59 for RCR, TSA and CMC I patients, respectively.

One year after surgery, the highest possible EQ-5D-5L utility index of 1.0 was found in 54%, 36% and 17% of the RCR, TSA and CMC I patients, respectively, indicating a ceiling effect. The prevalence of the highest possible score in the disease-specific questionnaires was 34% of the OSS

Table 6 Responsiveness shown by effect sizes of the outcome measures stratified by intervention group

	RCR	TSA	CMC I
EQ-5D-5L utility index	1.3	1.1	1.3
Oxford Shoulder Score	2.4		
SPADI		2.5	
Brief MHQ			2.2
QuickDASH		1.7	1.6
Constant Score	1.4*	2.4	
Abduction strength	0.6*	1.6	
Key pinch strength			0.1

RCR rotator cuff repair, *TSA* total shoulder arthroplasty, *CMC I* thumb carpometacarpal joint arthroplasty, *SPADI* Shoulder Pain and Disability Index, *Brief MHQ* brief Michigan Hand Outcomes Questionnaire, *QuickDASH* short version of the Disabilities of the Arm, Shoulder and Hand questionnaire

*Calculated using the 6-months follow-up

Table 7 Minimal important difference (MID) and minimal important change (MIC) of the EQ-5D-5L utility index calculated by different methods

	RCR	TSA	CMC I
MID	0.048	0.094	0.027
MIC—distribution	0.098	0.102	0.090
MIC—mean change	0.161	0.183	0.209
MIC—ROC	0.130	0.058	0.166

RCR rotator cuff repair, *TSA* total shoulder arthroplasty, *CMC I* thumb carpometacarpal joint arthroplasty, *ROC* receiver operating characteristic curve

in the RCR group, 9% for the SPADI in the TSA group and 17% for the brief MHQ in the CMC I group. No floor effects for the EQ-5D-5L utility index (0%) were found in any group at any time point.

For all three populations, the EQ-5D-5L utility index differed significantly between the patients in a PASS versus those who were not, which indicates the good discriminative ability of this QoL outcome measure ($p \le 0.001$, Table 8).

Discussion

The analysis of EQ-5D-5L utility index measurement properties revealed overall good construct validity, responsiveness and discriminative ability in patients after arthroscopic RCR, TSA and CMC I arthroplasty. However, the questionnaire shows considerable ceiling effects 1 year after surgery. The calculations of MID and MIC yielded inconsistent results with values varying from 0.027 to 0.209.

Table 8 Discriminative ability of the EQ-5D-5L utility index for comparing patients in an acceptable symptom state (PASS) at any follow-up and those who are not

	RCR		TSA		CMC I	
	n	EQ-5D-5L utility index, mean (SD)	n	EQ-5D-5L utility index, mean (SD)	n	EQ-5D-5L utility index, mean (SD)
PASS	274	0.947 (0.078)	253	0.909 (0.133)	227	0.899 (0.116)
no PASS	24	0.663 (0.217)	43	0.741 (0.164)	73	0.800 (0.159)
p-value		≤0.001		≤0.001		≤0.001

p-values report the results of the Mann–Whitney U test for differences in the EQ-5D-5L utility index for patients in a PASS or not

RCR rotator cuff repair, *TSA* total shoulder arthroplasty, *CMC I* thumb carpometacarpal joint arthroplasty, *n* number of patients

Validity

All three hypotheses for testing construct validity were confirmed. Correlations between EQ-5D-5L and the disease-specific questionnaires, which measure joint-specific functional outcomes, were moderate to high in our three groups. Previous reports presented moderate ($r = 0.57$) and high correlations ($r = 0.77$) for the OSS versus EQ-5D three-level questionnaire [35] and EQ-5D-5L [36], respectively, in patients with shoulder disorders including rotator cuff tears, which also indicates good construct validity. Although it is unclear which EQ-5D version was examined, Dabija and Jain [37] found a negative correlation between the EQ-5D and SPADI in patients with rotator cuff tears; their correlation coefficient of -0.66 is similar to our value for TSA patients.

The correlation between the EQ-5D-5L and hand-specific brief MHQ was slightly lower in comparison to QoL versus the shoulder scores, yet closely matches the correlation established for patients with Dupuytren's disease ($r = 0.60$) [10]. Further studies investigated correlations of the EQ-5D-5L and the original full MHQ and presented coefficients of 0.5 and 0.63 for patients with carpal tunnel syndrome [6] and various hand or wrist conditions [38], respectively. In general, the moderate correlations presented in this study and the aforementioned articles are due to the fact that the constructs measured by the EQ-5D-5L (i.e. quality of life) differ from the disease-specific patient-reported outcome measures (PROMs) that specifically focus on measuring shoulder or hand function.

Responsiveness

Responsiveness of the EQ-5D-5L was high for all three groups with effect sizes of 1.1 and 1.3. Consequently, our predefined hypothesis that the effect sizes of the EQ-5D-5L utility index would be greater than or equal to 0.8 in all three cohorts was confirmed. A similar effect size of 1.186 was reported for EQ-5D (version unknown) assessments of patients after TSA [8]. Evans et al. [9] and Olerud et al.

[4] also revealed high responsiveness of the EQ-5D-3L in patients after RCR, TSA and with a proximal humeral fracture. Although our study showed high responsiveness of the EQ-5D-5L after CMC I arthroplasty, this aspect was only low to moderate for other hand surgery interventions; the effect size was 0.25 in patients treated for Dupuytren's disease [10] and 0.5 for patients after carpal tunnel release [6].

Responsiveness of the EQ-5D-5L in our three cohorts was expectedly lower than that noted for other disease-specific PROMs, as confirmed by the working groups of Evans, Hensler, Jain and Marti [6, 8–10]. This indicates that disease-specific PROMs are more suitable for assessing the treatment effect in these specific patient groups and should be used as outcome measures when subjective joint function is of primary interest. However, QoL significantly increased for all our patients 1 year after surgery and the QoL scores exceeded the EQ-5D-5L norm values, which have been determined for different age groups and are 0.87 for RCR patients and 0.85 for TSA and CMC I patients [39].

Interpretability

The calculation of the MID and MIC yielded different values with large variances depending on the calculation method applied. Therefore, these inconsistent results should be considered with caution. A MIC of 0.09 has been calculated for patients after carpal tunnel release [6] and a simulation-based study revealed a MID of 0.083 for the German value set of the EQ-5D-5L independent of a specific disease or intervention [40]. Earlier published MIDs were 0.074 [41] or equivalent to half of a SD of the baseline score [30]. The latter study employed a distribution-based approach, which only considers the variance of the data and not the patient's perspective. If possible, anchor-based approaches are preferred for MID and MIC calculations, which include a definition of what is considered as "minimally important" by the patient [42]. We used multiple methods to draw a comprehensive picture of the MIC estimate, as has also been suggested by Crosby et al. [43].

In the literature, ceiling effects of the EQ-5D-5L utility index ranging from 16 to 57% have been reported [4–6, 9]. Furthermore, we found a considerable ceiling for the EQ-5D-5L utility index at follow-up, at least for the RCR and TSA patients. However, ceiling also occurred in the disease-specific questionnaires for the RCR and CMC I patients, indicating that the patients' health status 1 year after surgery is very good. Therefore, the high EQ-5D-5L utility indices can be partly attributed to the good health status of the patients and not to a ceiling effect in general. [4–6, 9].

Limitations

As this analysis of measurement properties is a secondary analysis of a clinical trial, our patients did not complete the questionnaire twice within a short time frame, which precludes the calculation of test–retest reliability and measurement error. Furthermore, MID and MIC values differed considerably for all three groups depending on the calculation method because only a few patients indicated that their QoL was unchanged or worse than before surgery. This situation resulted in insufficient AUC values, particularly for the TSA and CMC I groups, and indicates an unreliable MIC; the number of unimproved patients must be higher to draw reliable conclusions about the MID and MIC. To determine the MIC, we used only one retrospective anchor question, which might have introduced recall bias, as patients might not exactly remember the preoperative status of the previous year [44].

Conclusion

The EQ-5D-5L utility index shows good construct validity, responsiveness and discriminative ability in patients after arthroscopic RCR, TSA and CMC I arthroplasty. Despite the ceiling effects, the EQ-5D-5L seems to be a suitable tool for quantifying QoL in these patients, which is necessary to calculate quality-adjusted life years for cost-utility analyses.

Data availability Data will be made available upon request.

Code availability The code will be made available upon request.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

Ethical approval This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the Cantonal Ethics Committee of Zurich, Switzerland (no. 2013-0381).

Consent to participate Informed consent was obtained from all individual participants included in the study.

References

1. Narvani, A. A., Imam, M. A., Godeneche, A., Calvo, E., Corbett, S., Wallace, A. L., et al. (2020). Degenerative rotator cuff tear, repair or not repair? A review of current evidence. *Annals of the Royal College of Surgeons of England, 102*(4), 248–255

2. Carter, M. J., Mikuls, T. R., Nayak, S., Fehringer, E. V., & Michaud, K. (2012). Impact of total shoulder arthroplasty on generic and shoulder-specific health-related quality-of-life measures: A systematic literature review and meta-analysis. *Journal of Bone and Joint Surgery. American Volume, 94*(17), e127

3. Wajon, A., Vinycomb, T., Carr, E., Edmunds, I., & Ada, L. (2015). Surgery for thumb (trapeziometacarpal joint) osteoarthritis. *Cochrane Database of Systematic Reviews, 2*(2), CD004631

4. Olerud, P., Tidermark, J., Ponzer, S., Ahrengart, L., & Bergström, G. (2011). Responsiveness of the EQ-5D in patients with proximal humeral fractures. *Journal of Shoulder and Elbow Surgery, 20*(8), 1200–1206

5. Slobogean, G. P., Noonan, V. K., & O'Brien, P. J. (2010). The reliability and validity of the Disabilities of Arm, Shoulder, and Hand, EuroQol-5D, Health Utilities Index, and Short Form-6D outcome instruments in patients with proximal humeral fractures. *Journal of Shoulder and Elbow Surgery, 19*(3), 342–348

6. Marti, C., Hensler, S., Herren, D. B., Niedermann, K., & Marks, M. (2016). Measurement properties of the EuroQoL EQ-5D-5L to assess quality of life in patients undergoing carpal tunnel release. *Journal of Hand Surgery 41*(9), 957–962

7. Grobet, C., Marks, M., Tecklenburg, L., & Audigé, L. (2018). Application and measurement properties of EQ-5D to measure quality of life in patients with upper extremity orthopaedic

disorders: A systematic literature review. *Archives of Orthopaedic and Trauma Surgery, 138*(7), 953–961

8. Jain, S. S., DeFroda, S. F., Paxton, E. S., & Green, A. (2020). Patient-reported outcome measures and health-related quality-of-life scores of patients undergoing anatomic total shoulder arthroplasty. *Journal of Bone and Joint Surgery. American Volume, 102*(13), 1186–1193

9. Evans, J., Dattani, R., Ramasamy, V., & Patel, V. (2018). Responsiveness of the EQ-5D-3L in elective shoulder surgery: Does it adequately represent patient experience? *Journal of Orthopaedic Surgery (Hong Kong), 26*(2), 2309499018774922

10. Hensler, S., Wehrli, M., Herren, D. B., & Marks, M. (2020). Measurement properties of the German Unite Rhumatologique des Affections de la Main (URAM) scale in patients treated for Dupuytren's disease. *Hand Surgery and Rehabilitations, 39*(6), 568–574

11. Rundgren, J., Enocson, A., Mellstrand Navarro, C., & Bergstrom, G. (2018). Responsiveness of EQ-5D in patients with a distal radius fracture. *Hand (N Y), 13*(5), 572–580

12. Grobet, C., Audigé, L., Eichler, K., Meier, F., Brunner, B., Wieser, S., et al. (2020). Cost-utility analysis of arthroscopic rotator cuff repair: A prospective health economic study using real-world data. *Arthroscopy Sports Medicine and Rehabilitation, 2*(3), e193–e205

13. Grobet, C. E., Glanzmann, M. C., Eichler, K., Rickenbacher, D., Meier, F., Brunner, B., & Audigé, L. (2021) Cost-utility analysis of total shoulder arthroplasty: A prospective health economic study using real-world data. Journal of Shoulder and Elbow Surgery

14. von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gotzsche, P. C., & Vandenbroucke, J. P. (2007). The strengthening the reporting of observational studies in epidemiology (STROBE) statement: Guidelines for reporting observational studies. *Bulletin of the World Health Organization, 85*(11), 867–872

15. Herdman, M., Gudex, C., Lloyd, A., Janssen, M., Kind, P., Parkin, D., et al. (2011). Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Quality of Life Research, 20*(10), 1727–1736

16. Ludwig, K., Graf von der Schulenburg, J. M., & Greiner, W. (2018). German value set for the EQ-5D-5L. *PharmacoEconomics, 36*(6), 663–674

17. Dawson, J., Rogers, K., Fitzpatrick, R., & Carr, A. (2009). The Oxford shoulder score revisited. *Archives of Orthopaedic and Trauma Surgery, 129*(1), 119–123

18. Roach, K. E., Budiman-Mak, E., Songsiridej, N., & Lertratanakul, Y. (1991). Development of a shoulder pain and disability index. *Arthritis Care and Research, 4*(4), 143–149

19. Waljee, J. F., Kim, H. M., Burns, P. B., & Chung, K. C. (2011). Development of a brief, 12-item version of the Michigan Hand Questionnaire. *Plastic and Reconstructive Surgery, 128*(1), 208–220

20. Beaton, D. E., Wright, J. G., Katz, J. N., & Group, U. E. C. (2005). Development of the QuickDASH: Comparison of three item-reduction approaches. *Journal of Bone and Joint Surgery 87*(5), 1038–1046

21. de Vet, H. C. W., Terwee, C. B., Mokkink, L. B., & Knol, D. L. (2011). *Measurement in medicine: A practical guide*. Cambridge University Press.

22. Jaeschke, R., Singer, J., & Guyatt, G. H. (1989). Measurement of health status. Ascertaining the minimal clinically important difference. *Controlled Clinical Trials, 10*(4), 407–415

23. Marks, M., Hensler, S., Wehrli, M., Schindele, S., & Herren, D. B. (2019). Minimal important change and patient acceptable symptom state for patients after proximal interphalangeal joint arthroplasty. *Journal of Hand Surgery, 44*(2), 175–180

24. Constant, C. R., Gerber, C., Emery, R. J., Sojbjerg, J. O., Gohlke, F., & Boileau, P. (2008). A review of the Constant score:

Modifications and guidelines for its use. *Journal of Shoulder and Elbow Surgery, 17*(2), 355–361

25. Mokkink, L. B., Terwee, C. B., Knol, D. L., Stratford, P. W., Alonso, J., Patrick, D. L., et al. (2010). The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: A clarification of its content. *BMC Medical Research Methodology, 10*, 22

26. Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., et al. (2010). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology, 63*(7), 737–745

27. Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159

28. de Vet, H. C., Beckerman, H., Terwee, C. B., Terluin, B., & Bouter, L. M. (2006). Definition of clinical differences. *Journal of Rheumatology, 33*(2), 434

29. Copay, A. G., Subach, B. R., Glassman, S. D., Polly, D. W., Jr., & Schuler, T. C. (2007). Understanding the minimum clinically important difference: A review of concepts and methods. *The Spine Journal, 7*(5), 541–546

30. Norman, G. R., Sloan, J. A., & Wyrwich, K. W. (2003). Interpretation of changes in health-related quality of life: The remarkable universality of half a standard deviation. *Medical Care, 41*(5), 582–592

31. Terwee, C. B., Bot, S. D., de Boer, M. R., van der Windt, D. A., Knol, D. L., Dekker, J., et al. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology, 60*(1), 34–42

32. Tubach, F., Dougados, M., Falissard, B., Baron, G., Logeart, I., & Ravaud, P. (2006). Feeling good rather than feeling better matters more to patients. *Arthritis and Rheumatism, 55*(4), 526–530

33. Christie, A., Dagfinrud, H., Garratt, A. M., Ringen Osnes, H., & Hagen, K. B. (2011). Identification of shoulder-specific patient acceptable symptom state in patients with rheumatic diseases undergoing shoulder surgery. *Journal of Hand Therapy, 24*(1), 53–60

34. Marks, M., Grobet, C., Audigé, L., & Herren, D. B. (2019). Clinical thresholds of symptoms for deciding on surgery for trapeziometacarpal osteoarthritis. *Journal of Hand Surgery, 44*(9), 937–945

35. Zhaeentan, S., Legeby, M., Ahlstrom, S., Stark, A., & Salomonsson, B. (2016). A validation of the Swedish version of the WORC index in the assessment of patients treated by surgery for subacromial disease including rotator cuff syndrome. *BMC Musculoskeletal Disorders, 17*, 165

36. Haragus, H., Prejbeanu, R., Patrascu, J., Faur, C., Roman, M., Melinte, R., et al. (2018). Cross-cultural adaptation and validation of the Romanian Oxford Shoulder Score. *Medicine (Baltimore), 97*(23), e10926

37. Dabija, D. I., & Jain, N. B. (2019). Minimal clinically important difference of shoulder outcome measures and diagnoses: A systematic review. *American Journal of Physical Medicine and Rehabilitation, 98*(8), 671–676

38. Atthakomol, P., Manosroi, W., Sanguanrungsirikul, S., Punoppamas, S., Benjachaya, S., Tongprasert, S., et al. (2020). A Thai version of the Michigan hand questionnaire (Thai MHQ): An investigation of the psychometric properties. *Health and Quality of Life Outcomes, 18*(1), 313

39. Grochtdreis, T., Dams, J., Konig, H. H., & Konnopka, A. (2019). Health-related quality of life measured with the EQ-5D-5L: Estimation of normative index values based on a representative German population sample and value set. *The European Journal of Health Economics, 20*(6), 933–944

40. Henry, E. B., Barry, L. E., Hobbins, A. P., McClure, N. S., & O'Neill, C. (2020). Estimation of an instrument-defined minimally

important difference in EQ-5D-5L Index scores based on scoring algorithms derived using the EQ-VT version 2 valuation protocols. *Value in Health, 23*(7), 936–944

41. Walters, S. J., & Brazier, J. E. (2005). Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. *Quality of Life Research, 14*(6), 1523–1532

42. de Vet, H. C., Terwee, C. B., Ostelo, R. W., Beckerman, H., Knol, D. L., & Bouter, L. M. (2006). Minimal changes in health status questionnaires: Distinction between minimally detectable change and minimally important change. *Health and Quality of Life Outcomes, 4*, 54

43. Crosby, R. D., Kolotkin, R. L., & Williams, G. R. (2003). Defining clinically meaningful change in health-related quality of life. *Journal of Clinical Epidemiology, 56*(5), 395–407

44. de Vet, H. C., Terluin, B., Knol, D. L., Roorda, L. D., Mokkink, L. B., Ostelo, R. W., et al. (2010). Three ways to quantify uncertainty in individually applied "minimally important change" values. *Journal of Clinical Epidemiology, 63*(1), 37–45

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.