BMC Musculoskeletal Disorders

RESEARCH

Open Access

A systematic review of patient-reported outcome measures for idiopathic adhesive capsulitis - recommendations based on analyses of 16 existing questionnaires

Gustav K. M. Vangsgaard[1*], Michael R. Krogsgaard[1] and Christian F. Hansen[1]

Abstract

Background Patient reported outcome measures (PROMs) are essential to measure the patients' perspective in clinical studies. Like any measurement instrument, a PROM should be valid, reliable, and responsive. Adequate content validity relies on patient input, while construct validity can only be ensured by Modern Test Theory (MTT) models. Inadequate PROMs induce a significant risk of measurement errors. Currently, there is no thorough analysis of PROMs used in clinical research on idiopathic adhesive capsulitis (AC). The aim was to identify all PROMs used to evaluate AC, analyze their content and construct validity, and summarize the results in a recommendation on which PROM is the best for trials on AC.

Methods Musculoskeletal PROMs used to evaluate patients with AC were identified through PubMed searches in November 2024. Development and validity studies were identified for each PROM. Content validity was assessed based on existing guidelines emphasizing the involvement of patients in the development. Construct validity was assessed based on existing guidelines emphasizing the use of MTT models in the analysis. Both content and construct validity were rated from 1 to 5, and a concluding, aggregated assessment was made.

Results 16 different PROMs, used up to 45 times, were identified. 79 articles on measurement properties were identified and analyzed. None of the PROMs had been developed specifically for patients with AC. Four PROMs were developed by the involvement of patients but with other conditions than AC. Five PROMs had been validated with an MTT model. However, all five possessed inadequate content validity. Hence, it was not possible to identify any PROM with adequate content and construct validity for patients with AC.

Conclusion An adequate PROM for idiopathic AC does not exist and a new condition-specific questionnaire is needed. The existing PROMs should be used with significant reservations and results obtained by these PROMs should be interpreted with caution.

Keywords Patient-reported outcome measures, PROM, Frozen shoulder, Adhesive capsulitis

*Correspondence:
Gustav K. M. Vangsgaard
kalle2312@gmail.com

[1]Section of Sports Traumatology M51, Department of Orthopedic Surgery, Copenhagen University Hospital Bispebjerg and Frederiksberg, Gustav Kalle Mølbak Vangsgaard, Bispebjerg Bakke 23, 5, 1, Copenhagen, Copenhagen 2400, NV, Denmark

Introduction

Idiopathic adhesive capsulitis (AC), often referred to as 'frozen shoulder', causes substantial pain and reduced range of motion [1]. It is located specifically in the shoulder joint capsule and not in the other collagen tissues of the shoulder, and it is caused by a marked increase in production of all the types of collagen present in the joint capsule [2]. It is more common in persons with diabetes, Dupuytren's contracture and thyroid hypofunction [1], but most patients have no co-morbidities [1]. The condition is common with a lifetime prevalence of 2–5% [1, 3]. Fractures, development of scar-tissue following a trauma, or an operation and other shoulder diseases can also result in reduced range of motion and severe shoulder pain and is sometimes referred to as secondary frozen shoulder. However, this condition is not caused by the frozen shoulder process and is not subject for the current study.

There is a variety of treatments for AC, such as oral corticosteroid, local corticosteroid injection, physiotherapy, hydrodistension, manipulation under anesthesia, arthroscopic capsular release, and open capsular release [4], and no consensus on which is most efficient [5]. The treatment effect can be evaluated in different ways, and common measures are active and passive range of motion, function during loaded and everyday activities, and patient reported outcome measures (PROMs). The latter is increasingly regarded as highly important due to the ability to measure the patients' perspective without the interpretation of a physician [6, 7].

In research comparing groups, e.g., randomized controlled treatment studies, it is essential to use a valid PROM that can adequately measure the effect of treatments [4]. Studies using an adequate PROM as outcome find more than twice as often a significant difference between treatments than studies using a suboptimal PROM [8], and to use an inadequate PROM carries a high risk of a type 2 error in the study results. For a PROM to be considered adequate to evaluate AC, it must be thoroughly developed by the involvement of patients with AC, ensuring relevance and coverage of the content (i.e., content validity) [9–11], which is logical and also acknowledged by the COSMIN guidelines as the most important property of a PROM [9]. Its measurement properties are best psychometrically validated with a Modern Test Theory (MTT) model (e.g. Rasch Analysis or Confirmatory Factor Analysis (CFA), which are also termed Item Response Theory models), as these models are the strongest group of mathematical tools to assess the structure of the PROM and measurement properties of the items. These methods are used to exclude items of no value or with bad measurement properties, to identify domains, to identify differential item functioning in various groups of respondents and much more [8, 12]. Several other aspects of measurement, like responsiveness, are also important. This ensures that the PROM is accurately measuring the proposed construct (e.g., disability or pain) [13], and the risk of measurement error is minimized.

The only thoroughly structured, large catalogue on the content and construct validity of PROMs used in musculoskeletal research [12] includes nine PROMs that had at least once been validated for patients with AC. However, this catalogue did not specifically focus on PROMs for patients with AC, and relevant PROMs may not have been identified. Also, this catalogue does not offer any recommendation on which PROM is the most adequate to evaluate patients with AC.

The research question for this systematic review was to identify the most adequate PROMs to evaluate treatment outcome in patients with AC.

This is achieved by, (i) identification of all musculoskeletal PROMs that have been used as an outcome in studies evaluating AC, (ii) evaluation of the content validity of these PROMs, through an analysis of the development, (iii) evaluation of all published analyses of the measurement properties (i.e., criterion and construct validity, reliability, and responsiveness) of these PROMs, and (iv) by summarizing the results in a recommendation. PROMs that are not intended specifically for musculoskeletal conditions (e.g., generic PROMs like EQ-5D and SF-36) and single scores (e.g., VAS for pain) were not included.

Materials and methods

The study followed the PRISMA guidelines for reporting of systematic reviews [14].

Inclusion and exclusion criteria for the database search

Condition-specific PROMs that have been used in clinical studies evaluating patients with AC were included. Generic PROMs and single scores were excluded. Studies that did not specifically mention patients with AC (but for instance only "shoulder conditions") and studies that did not specify which PROM had been used were excluded. Articles in other languages than English or Scandinavian languages were excluded.

Identification of relevant PROMs

Relevant PROMs were identified November 27th 2024 through searches in PubMed with the search string: ((frozen shoulder) OR (capsulitis)) AND (questionnaire OR PROM OR "patient reported outcome" OR "patient-reported outcome") AND (*measurement filters*) NOT (*exclusion filter*) with measurement filters and exclusion filters being algorithms developed for the purpose of identifying PROMs with a sensitivity of 97% [15, 16]. The studies were manually screened by the first author on

title, abstract, and full text when relevant. Uncertainties were discussed in the group of authors.

Also, the validation studies in the aforementioned catalogue of musculoskeletal PROMs were screened manually, and studies that had included patients with AC were identified [12]. In case the PROM had not be identified in the search in PubMed (because it had never been used in clinical research on patients with AC), it was also included, since it could have adequate measurement properties for the patient group.

Identification of the development study for each PROM

For the assessment of content validity, studies describing the original developmental process were identified for each PROM through simple searches on the name of the PROM in PubMed.

Identification of validity studies for the selected PROMs

Similarly, studies evaluating the measurement properties (i.e., criterion and construct validity, reliability, responsiveness) and studies reporting translations and intercultural adaptations of each PROM were identified through subsequent searches in PubMed November 28th 2024: ("PROM name") AND (validity OR validation OR adaption OR adaptation OR translation) AND (*measurement filters*) NOT (*exclusion filter*). The studies were manually screened by the first author on title, and by abstract and full text when relevant. Studies on measurement properties were included if they involved patients with AC. In cases where the diagnoses of the included shoulder patients were not specified (e.g., 'unspecific shoulder pain' or 'unspecified conditions of the shoulder'), the study was included. English and Scandinavian language studies only were included.

Quality assessment

Assessment of the quality of the development and the validation of the measurement properties of the PROMs were performed according to previously published guidelines [12, 15].

The quality of a PROM depends mainly on its development, as this secures content validity (relevance and coverage) [9], and secondarily on the measurement properties, which include criterion and construct validity, reliability, and responsiveness [9].

All included studies were assessed by two authors independently, and ratings were then discussed until agreement. If the two authors were unable to agree on a final rating, the third author was consulted.

Quality assessment of the development

Content validity is the most important property of a PROM [9–11] since it defines what the PROM is measuring. The content must be based on input from patients of the target population [11]. This is obtained through a qualitative approach, typically semi-structured group interviews, to discuss relevant themes and items (i.e., questions and response options) until no further themes emerge [9]. Detailed guidelines on how to develop condition-specific PROMs are available [9–11].

This study utilizes a quality assessment method described in 2021 [12]. The method was developed in a study group consisting of clinicians and statisticians with special expertise on theoretical and practical aspects of PROMs. The method was used for work on the large catalogue of musculoskeletal PROMs [12] before the detailed COSMIN guidelines were published [17]. To evaluate whether the conclusions from the two analytic methods (the current method and COSMIN's guideline) were different, an analysis after COSMIN's guidelines was performed for five different PROMs [18], and there were no important differences between the two assessments. As the COSMIN guidelines are quite time-consuming to use, we have chosen for the current study to use the system we developed.

The development and validity studies of the identified PROMs were methodologically rated from 1 to 5, 1 being the lowest quality and 5 being the highest. For development studies, a rating of 1 was given if the items had not been generated with the involvement of patients or experts and the PROM did not include items from previously known PROMs or item banks. From here, the rating was increased by one point for each of the following: (i) the PROM was based on items generated by experts, or on previously known PROMs or item banks, (ii) the items were based on interviews with a well targeted group of patients, (iii) interviews conducted were continued until no further items emerged (i.e. data saturation), and (iv) the content had been debriefed with a targeted group of patients in its final form. Thus, the maximum rating of a development study that had not involved any patients was 2 out of 5. A rating of at least 4 is required for the content validity of a PROM to be regarded as acceptable.

Quality assessment of the construct validation

A developmental process produces a preliminary PROM, and, subsequently, its psychometric properties must be assessed. This process secures the structure of the PROM (typically in several domains) and removes items that do not contribute meaningful information to the scores. Based on this, the final version of the PROM emerges. This ensures the ability of the PROM to accurately measure the proposed construct (i.e., pain) validly and reliably. The studies assessing psychometric properties were rated from 1 to 5, with 1 being the lowest quality and 5 being the highest. There is substantial evidence that modern test theory/items response theory models, such as Rasch Analysis or CFA best evaluate the structure and

measurement properties of a PROM, compared to classical test theory methods [13, 19]. MTT-models offer analyses of all relevant properties related to validity, including several imperative properties (e.g., unidimensionality, differential item functioning, and local dependence) that are simply assumed (although often not present) with CTT methods [13]. There are different strengths and limitations to the various MTT models, but they were regarded as equal in the current assessment [13]. Criterion validity relates to how a PROM correlates with a 'gold standard'. However, since there is no 'gold standard' for most PROMs, it often concerns how the PROM in question correlates to a similar PROM, Visual Analogue Pain, or Numeric Rating of Pain, and whether the correlations fit a predefined hypothesis/expectation [13]. There are different types of reliability, but in this assessment, it was acceptable if test-retest, intraclass correlation, or calculation of Cronbach's alpha had been reported. Responsiveness concerns whether a PROM can measure a true change over time.

The validity studies were rated with points as follows: (i) if a study validated a condition with less than 20 patients included in the analyses, the methodological rating would be 1 regardless of the methods used. A study would also be rated 1 if no actual statistical analyses were undertaken (e.g., translation only), (ii) if validity, reliability, and responsiveness had all been assessed, but an MTT model analysis had not been used, the study was rated 3, (iii) all scenarios between i and ii were rated 2 (e.g., no MTT model used and only assessment of validity and reliability, but not responsiveness), (iv) if any MTT model had been used (e.g. Rasch or CFA) the study would be rated 4. If, in addition, reliability and responsiveness had also been assessed, the study was rated 5. There are important structure- and measurement-properties that can only be assessed through MTT model analyses, and PROMs that have only been assessed by classical test theory methods cannot be considered to have confirmed adequate measurement properties [10].

Quality of the PROM (aggregated assessment)

Since a PROM needs to have content as well as construct validity, the aggregated score can only be as high as the lowest rating of the two. In addition, the result of the construct validation is also considered, meaning than if it shows that the measurement properties are inadequate, the aggregated score will be corrected accordingly (e.g., a study with a Rasch analysis where the data exhibited *bad fit* to the model would not receive points for the validation with an MTT analysis, but only according to the remaining validation in the study).

Results

The initial search to identify all PROMs relevant for AC resulted in 364 studies. With addition of studies from meta-analyses it yielded 133 studies on AC, in which 39 different PROMs were used as outcomes at least once (Fig. 1). Of these, 25 PROMs were excluded since they were not musculoskeletal, were generic questionnaires, simple scales and not PROMs, or it was not possible to identify which PROM the study had used (Fig. 1). This led to the inclusion of 14 PROMs, used between 1 and 45 times (Table 1). The total count used (175) exceeds the number of included studies (133) since some studies had more than one PROM as outcome. Two additional PROMs, known to be previously validated for patients with AC, but which had never been used in a study on AC, were included from the catalogue [12].

The searches related to studies investigating the measurement properties of the included PROMs yielded 1662 articles, from which 39 were found relevant, and included for assessment. Adding further articles from the catalogue [12], a total of 79 validation studies were included for assessment (Fig. 2).

The assessments related to the development of the 16 PROMs are presented in Supplementary Table 1. The assessments of the 79 validity studies are presented in Supplementary Table 2.

An overview in relation to the 133 records that were reviewed in full text is presented in Supplementary Table 3.

There was no PROM developed specifically for AC. Patients had been involved in the item generation of four of the 16 PROMs: Croft Disability Questionnaire (Croft) [20], Flexilevel Scale of Shoulder Function (FLEX-SF) [21], Oxford Shoulder Score (OSS) [22], and Shoulder Disability Questionnaire (SDQ) [23]. These four PROMs measure the disability associated with shoulder symptoms, shoulder function, surgically treatable shoulder conditions, and functional limitations in patients with soft tissue shoulder disorders, respectively (Supplementary Table 1). However, items had not been discussed until data saturation during the development process in any of these four PROMs. OSS debriefed the items with patients, and its development was rated 4, while the other three were rated 3. The remaining 12 PROMs did not involve patients in the generation of items and their development was rated 2 or less (Table 2, Supplementary Table 1).

Regarding the construct validation of the PROMs (Supplementary Table 2), five PROMs had been analyzed with an MTT model and were rated at least 4, while one (Quick Disability of the Arm, Shoulder, and Hand (Q-DASH) [24]) achieved the maximum score of 5 (Table 2).

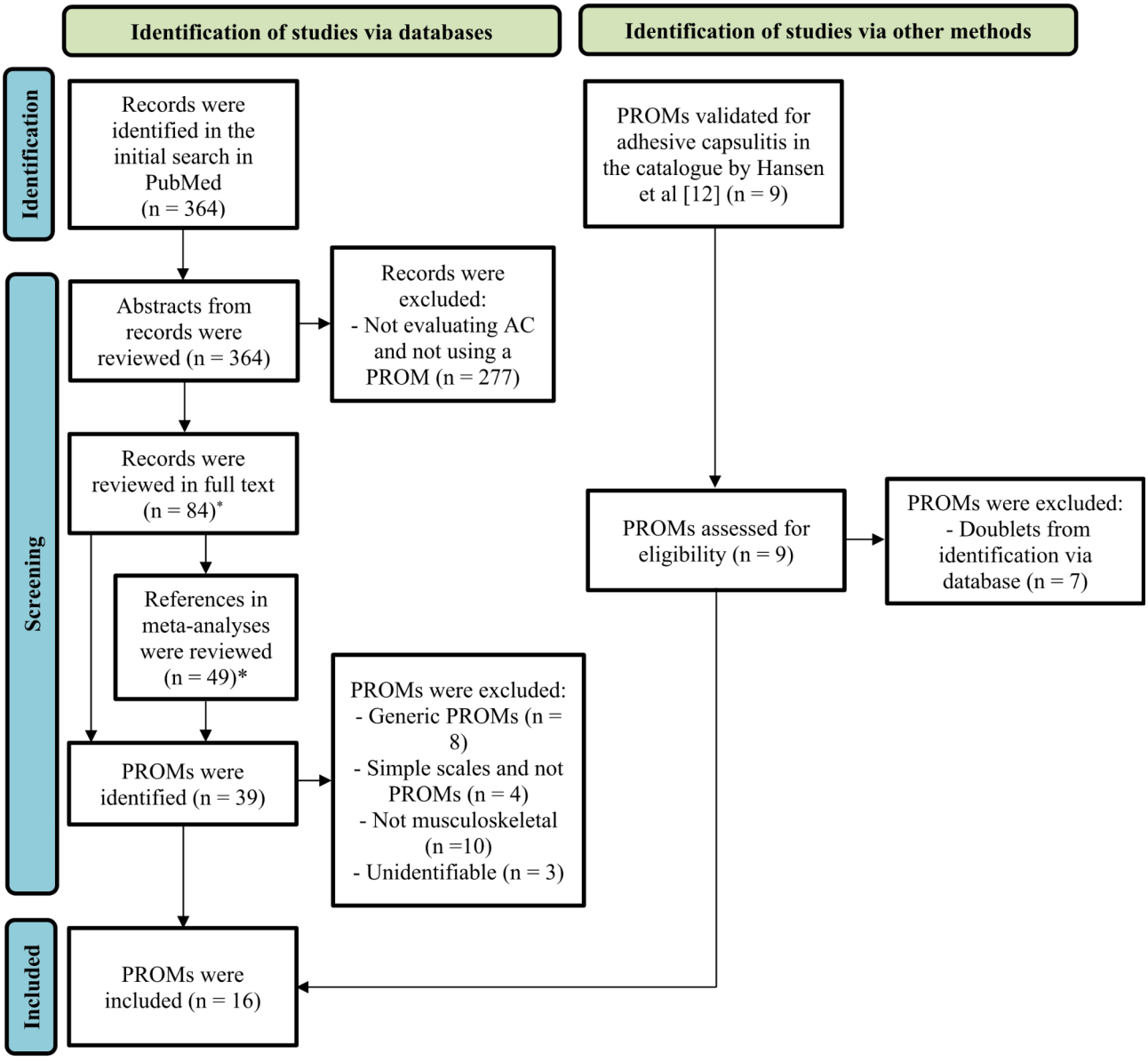

Fig. 1 Flow chart with numbers of the included and excluded records and PROMs. AC; adhesive capsulitis, PROM; patient-reported outcome measures. *: An overview in relation to the 133 records that were reviewed in full text is presented in Supplementary Table 3

Ten of the 16 PROMs were psychometrically validated in a cohort of patients with AC (Supplementary Table 2). One of these had been analyzed with an MTT model, thus rated 4 (Shoulder Pain and Disability Index (SPADI) [25]). The psychometric validation of the remaining 9 PROMs were all rated 3 or less (Table 2).

In the aggregated assessment (Table 2), two PROMs (FLEX-SF and SDQ) were rated 3, while the rest were rated less. However, neither of the two targeted patients with AC.

Discussion

Based on this systematic review, there is no PROM with adequate content validity for patients with AC. Four PROMs (Croft, FLEX-SF, OSS, SDQ) involved patients in their development process, but no one had included patients with AC. Thus, these four PROMs have some degree of content validity, but for broader or different patient groups with shoulder problems. It is questionable if their content is valid for patients with AC and thus questionable what they measure when used in this patient group. No PROM was rated the maximum score of 5, as interviews with patients had not been continued until no further items emerged (i.e., to data saturation),

Table 1 The total number of times a PROM is used in the clinical research of adhesive capsulitis and their development year. Sorted after the number of times used

Number of times used	PROM	Development year
45	Shoulder Pain and Disability Index (SPADI)	1991
37	Constant-Murley Score (CMS)	1987
20	American Shoulder and Elbow Surgeons Score (ASES)	1999
18	Disabilities of the Arm, Shoulder and Hand (DASH)	1996
14	Shoulder Disability Questionnaire (SDQ)	2000
13	Oxford Shoulder Score (OSS)	1996
9	Simple shoulder test (SST)	1992
6	University of California Los Angeles Shoulder Scale (UCLASS)	1981
4	Shoulder Rating Questionnaire (SRQ)	1997
2	Quick Disabilities of the Arm, Shoulder and Hand (Q-DASH)	2005
2	The Upper Extremity Functional Index (UEFI)	2001
2	Flexilevel Scale of Shoulder Function (FLEX-SF)	2003
2	Croft Shoulder Disability Questionnaire (Croft)	1994
1	Rowe Score (Rowe)	1978
0	Single Assessment Numeric Evaluation (SANE)	1999
0	Patient-Reported Outcomes Measurement Information System Upper Extremities (PROMIS UE)	2013

PROM; patient-reported outcome measure

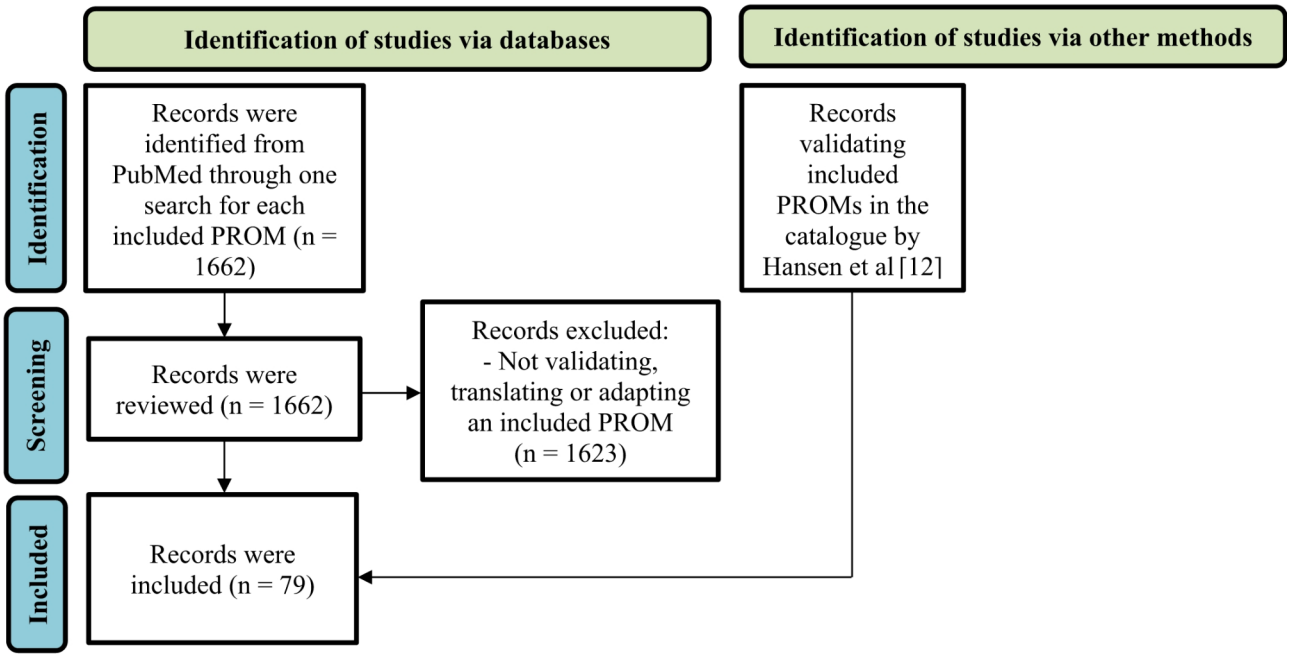

Fig. 2 Flow chart with numbers of the included and excluded records with psychometric PROM validation. PROM; patient-reported outcome measures.

which threatens sufficient coverage [11]. Croft, OSS, and SDQ have been psychometrically validated in cohorts of patients with AC. Croft and OSS were rated 2 and SDQ 3, since they have not been analyzed with an MTT model. Therefore, there is no indication that these four PROMs are adequate measurement instruments for patients with AC, even though they were developed with the involvement of patients.

Content validity is regarded as the most important quality of a PROM, and construct validity as less important, because it can potentially be improved any time by use of MTT analyses [9, 11]. Regarding the construct validity, Q-DASH scored the maximum points of 5 as it had been validated for patients with unspecified shoulder pain. However, the development did not include patient input and the PROM may not assess key aspects of the condition in patients with AC [26].

SPADI was the only PROM that was construct validated with an MTT model in a cohort of patients with AC, and with good fit to the model. Still, SPADI was

Table 2 Rating of content- and construct-validity for the 16 PROMs used or validated for patients with adhesive capsulitis

PROM	Development rating	Validity rating for 'shoulder disorder'	Validity rating for adhesive capsulitis	Aggregated score
ASES	2	4	-	2
CMS	1	2	-	1
Croft	3	0	2	2*
DASH	2	3	2	2
FLEX-SF	3	3	-	3
OSS	4	-	2	2*
PROMIS UE	2	4	2	2
Q-DASH	2	5	3	2
ROWE	1	-	-	0*
SANE	1	0	3	1
SDQ	3	2	3	3
SPADI	2	3	4	2
SRQ	2	2	3	2
SST	1	2	2	1
UCLASS	2	2	-	2
UEFI	2	4	-	2

PROM; patient-reported outcome measures, AC; adhesive capsulitis, ASES; American Shoulder and Elbow Surgeons Score, CMS; Constant-Murley Score, Croft; Croft Shoulder Disability Questionnaire, DASH; Disabilities of the Arm, Shoulder and Hand, FLEX-SF; Flexilevel Scale of Shoulder Function, OSS; Oxford Shoulder Score, PROMIS UE; Patient-Reported Outcomes Measurement Information System Upper Extremities, Q-DASH; Quick Disabilities of the Arm, Shoulder and Hand, Rowe: Rowe Score, SANE: Single Alpha Numeric Evaluation, SDQ; Shoulder Disability Questionnaire, SPADI; Shoulder Pain and Disability Index, SRQ; Shoulder Rating Questionnaire, SST; Simple Shoulder Test, UCLASS; University of California Los Angeles Shoulder Scale, UEFI; The Upper Extremity Functional Index

-; not assessed,

*; higher score possible if a better psychometric validation is performed

The score given for development, the highest score given in validation for unspecified shoulder condition and specifically for adhesive capsulitis, respectively, and the aggregated score for PROMs that has been used to evaluate adhesive capsulitis sorted alphabetically

developed for shoulder pathology in general, and patients had not been involved in the development. Therefore, it cannot be characterized as an adequate PROM for patients with AC.

It was not possible to identify any PROM with adequate content and construct validity for patients with AC. None of the four PROMs that were developed by the involvement of patients (Croft, FLEX-SF, OSS and SDQ) had been developed specifically for patients with AC and none of them have been construct validated with an MTT model. SPADI has good construct validity for patients with AC, so despite its insufficient development and content validity, it might be pragmatic to use it for patients with AC. However, to use it carries a high risk of low responsiveness, low specificity and type-II measurement errors [8, 19, 27]. As a consequence, it is recommended that a new, condition-specific PROM for patients with AC is developed according to modern-day guidelines [9]. Items and domains for a PROM which can be characterized as valid and adequate for patients with AC would be developed through group interviews with patients who have or have had AC, until no further topics are brought up (data saturation), with understandability of the wording evaluated through individual interviews, with construct optimized through the use of MTT methods in the analysis of about 150 completed provisional questionnaires and with responsiveness, reliability and minimal clinical important difference established through longitudinal studies with repeated completion of the questionnaire. As AC is a common condition such a specific PROM can be developed within a reasonable time frame.

The most frequently used PROM was SPADI with 45 hits. SPADI, Croft, FLEX-SF, OSS and SDQ seemed to be slightly better than the rest due to either content or construct validity but were used in less than half of all cases (76 of 175 times). The second most used PROM was Constant-Murley Score (CMS) (37 of 175), which achieved the lowest possible rating of 1 (Table 2), and scores obtained with CMS must be interpreted with significant reservation.

Obviously, the most frequently used PROMs are not the most valid. This is surprising since PROMs are scientific measurement instruments. It could be that the research community is not aware of the necessity for both content and construct validity of a PROM, or perhaps the potential consequences of using inadequate PROMs are overlooked. For most PROMs it is easy to get the impression that it is in fact valid because there are reviews and studies that conclude they are, even though the PROM does not live up to basic principles of validity when it is evaluated after strict principles [19]. Even though it may take years for a new PROM to be developed, recognized and generally accepted, it is imperative to develop valid outcome measures, as this minimizes the risk of measurement errors. Adequate PROMs are more responsive

and when they are used as outcomes it maximizes the chance of detecting a true difference between treatment groups [8, 27]. Some clinical treatment trials have a large impact on the daily clinic and the choice of the most appropriate treatment strategy, and if a true difference cannot be detected in such trials, because an invalid/inadequate PROM has been used as outcome, there is a risk that health care funds are used sub-optimally and that patients do not receive the best treatment. If no adequate PROM exists, it can be considered to choose a different outcome as the primary study outcome.

It is a limitation that the searches were only performed in one database (PubMed). The search string was specific for studies that had used a PROM for patients with AC, but the combination of our earlier and broader search in relation to our earlier catalogue [12] only identified two additional PROMs that have never been used for patients with AC. There is potentially publication bias, as studies showing inadequacy of a PROM are more difficult to publish that studies that find the PROM valid and useful–this stresses that a critical review of validation studies is necessary before conclusions on adequacy of a PROM can be made, and unfortunately most reviews on PROMs do not go in depth with this, meaning that scientifically unjustified positive evaluations are common.

The analyses were restricted to manuscripts in English or Scandinavian language, which might have excluded some studies on validation of the construct validity.

Conclusion

There is no PROM with adequate content and construct validity for patients with AC. In the attempt to determine the optimal treatment for AC a new and condition-specific PROM for AC is needed. The current PROMs should be used with significant reservation and scores obtained by them should be interpreted with caution.

Abbreviations

AC	Adhesive capsulitis
PROM	Patient reported outcome measure
MTT	Modern Test Theory
CFA	Confirmatory Factor Analysis
CTT	Classical Test Theory
Croft	Croft Disability Questionnaire
FLEX-SF	Flexilevel Scale of Shoulder Function
OSS	Oxford Shoulder Score
SDQ	Shoulder Disability Questionnaire
Q-DASH	Quick Disability of the Arm, Shoulder, and Hand
SPADI	Shoulder Pain and Disability Index
CMS	Constant-Murley Score

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12891-025-08443-z.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Acknowledgements
Not applicable.
The protocol for the study has not been registered.

Author contributions
GV did the searches and analyzed, together with CH, the data. GV wrote the main manuscript and made the tables and figures. All authors (GV, MK and CH) read, reviewed and approved the final manuscript.

Funding
Open access funding provided by Copenhagen University
The authors did not receive any funding.

Data availability
No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Conflict of interest
None.

Received: 16 July 2024 / Accepted: 17 February 2025
Published online: 09 May 2025

References

1. Sarasua SM, Floyd S, Bridges WC, Pill SG. The epidemiology and etiology of adhesive capsulitis in the U.S. Medicare population. BMC Musculoskelet Disord. 2021;22(1):828.
2. Marker L, Schjerling P, Mackey AL, Hansen T, Jakobsen J, Kjaer M, Krogsgaard MR. Collagens in primary frozen shoulder: expression of collagen mRNA isoforms in the different phases of the disease. Rheumatology (Oxford). 2021;60(8):3879–87.
3. Jacob L, Gyasi RM, Koyanagi A, Haro JM, Smith L, Kostev K. Prevalence of and risk factors for adhesive capsulitis of the shoulder in older adults from Germany. J Clin Med 2023, 12(2).
4. Cho CH, Bae KC, Kim DH. Treatment strategy for frozen shoulder. Clin Orthop Surg. 2019;11(3):249–57.
5. Xiao Y, Tang H, Meng J, Wu Y, Liu W, Liu P, Gao S. Similar outcomes between arthroscopic capsular release and manipulation under anaesthesia for frozen shoulder: A meta-analysis. Asian J Surg 2024.
6. Hildon Z, Neuburger J, Allwood D, van der Meulen J, Black N. Clinicians' and patients' views of metrics of change derived from patient reported outcome measures (PROMs) for comparing providers' performance of surgery. BMC Health Serv Res. 2012;12:171.
7. Patrick DL, Burke LB, Powers JH, Scott JA, Rock EP, Dawisha S, O'Neill R, Kennedy DL. Patient-reported outcomes to support medical product labeling claims: FDA perspective. Value Health. 2007;10(Suppl 2):S125–137.
8. Hansen CF, Jensen J, Brodersen J, Siersma V, Comins JD, Krogsgaard MR. Are adequate proms used as outcomes in randomized controlled trials? An Analysis of 54 trials. Scand J Med Sci Sports. 2021;31(5):972–81.
9. Comins JD, Brodersen J, Siersma V, Jensen J, Hansen CF, Krogsgaard MR. How to develop a condition-specific PROM. Scand J Med Sci Sports. 2021;31(6):1216–24.
10. Patrick DL, Burke LB, Gwaltney CJ, Leidy NK, Martin ML, Molsen E, Ring L. Content validity–establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product

evaluation: ISPOR PRO good research practices task force report: part 1–eliciting concepts for a new PRO instrument. Value Health. 2011;14(8):967–77.
11. Terwee CB, Prinsen CAC, Chiarotto A, Westerman MJ, Patrick DL, Alonso J, Bouter LM, de Vet HCW, Mokkink LB. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. Qual Life Res. 2018;27(5):1159–70.
12. Hansen CF, Jensen J, Siersma V, Brodersen J, Comins JD, Krogsgaard MR. A catalogue of proms in sports science: quality assessment of PROM development and validation. Scand J Med Sci Sports. 2021;31(5):991–8.
13. Christensen KB, Comins JD, Krogsgaard MR, Brodersen J, Jensen J, Hansen CF, Kreiner S. Psychometric validation of PROM instruments. Scand J Med Sci Sports. 2021;31(6):1225–38.
14. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71.
15. Hansen CF, Obionu KC, Comins JD, Krogsgaard MR. Patient reported outcome measures for ankle instability. An analysis of 17 existing questionnaires. Foot Ankle Surg. 2022;28(3):288–93.
16. Terwee CB, Jansma EP, Riphagen II, de Vet HC. Development of a methodological pubmed search filter for finding studies on measurement properties of measurement instruments. Qual Life Res. 2009;18(8):1115–23.
17. Gagnier JJ, Lai J, Mokkink LB, Terwee CB. COSMIN reporting guideline for studies on measurement properties of patient-reported outcome measures. Qual Life Res. 2021;30(8):2197–218.
18. Hansen CF, Jensen J, Odgaard A, Siersma V, Comins JD, Brodersen J, Krogsgaard MR. Four of five frequently used orthopedic proms possess inadequate content validity: a COSMIN evaluation of the mHHS, HAGOS, IKDC-SKF, KOOS and KNEES-ACL. Knee Surg Sports Traumatol Arthrosc. 2022;30(11):3602–15.
19. Krogsgaard MR, Hansen CF. Patient-reported outcome measures: it is time for authors, reviewers, journal editors and health care strategists to take sufficient responsibility. Knee Surg Sports Traumatol Arthrosc. 2022;30(11):3589–93.
20. Croft P, Pope D, Zonca M, O'Neill T, Silman A. Measurement of shoulder related disability: results of a validation study. Ann Rheum Dis. 1994;53(8):525–8.
21. Cook KF, Roddey TS, Gartsman GM, Olson SL. Development and psychometric evaluation of the flexilevel scale of shoulder function. Med Care. 2003;41(7):823–35.
22. Dawson J, Fitzpatrick R, Carr A. Questionnaire on the perceptions of patients about shoulder surgery. J Bone Joint Surg Br. 1996;78(4):593–600.
23. van der Heijden GJ, Leffers P, Bouter LM. Shoulder disability questionnaire design and responsiveness of a functional status measure. J Clin Epidemiol. 2000;53(1):29–38.
24. Beaton DE, Wright JG, Katz JN, Upper Extremity Collaborative G. Development of the QuickDASH: comparison of three item-reduction approaches. J Bone Joint Surg Am. 2005;87(5):1038–46.
25. Roach KE, Budiman-Mak E, Songsiridej N, Lertratanakul Y. Development of a shoulder pain and disability index. Arthritis Care Res. 1991;4(4):143–9.
26. Brodersen JDL, Thorsen H, Mckenna SP. Rasch Models in Health: Wiley; 2012.
27. Comins JD, Siersma VD, Lind M, Jakobsen BW, Krogsgaard MR. KNEES-ACL has superior responsiveness compared to the most commonly used patient-reported outcome measures for anterior cruciate ligament injury. Knee Surg Sports Traumatol Arthrosc. 2018;26(8):2438–46.

Publisher's note