Received: 25 May 2020 | Revised: 11 December 2020 | Accepted: 4 January 2021

DOI: 10.1111/sms.13923

ORIGINAL ARTICLE

WILEY

A catalogue of PROMs in sports science: Quality assessment of PROM development and validation

Christian F. Hansen[1] | **Jonas Jensen**[1] | **Volkert Siersma**[2] | **John Brodersen**[2,3] | **Jonathan D. Comins**[1,2] | **Michael R. Krogsgaard**[1]

[1]Section for Sports Traumatology M51, Bispebjerg and Frederiksberg Hospital, Copenhagen, Denmark

[2]The Research Unit for General Practice and Section of General Practice, Department of Public Health, University of Copenhagen, Copenhagen, Denmark

[3]Primary Health Care Research Unit, Region Zealand, Sorø, Denmark

Correspondence
Michael R. Krogsgaard, Section for Sports Traumatology M51, Bispebjerg and Frederiksberg Hospital, Copenhagen, Denmark.
Email: michael.rindom.krogsgaard@regionh.dk

Choosing the most adequate PROM for a study is a non-trivial process. The aim of this study was to provide a catalogue with analyses of content and construct validity of PROMs relevant to research in sports science, including all published local translations. The most commonly used PROMs in sports research were selected from a PubMed search "patient reported outcome measures sports", identifying 439 articles and 194 different PROMs. Articles describing development of the 61 selected PROMs were assessed for content validity, and all articles regarding construct validity of each PROM and all published translations (in total 622 articles) were analyzed. A catalogue with assessments of the 61 PROMs was produced. The majority were of inferior validity, with few exceptions. The most common reason for this was that the PROM had not been developed by methods that ensure high content validity. Another major reason for inferior validity was that construct validity had not been secured by adequate statistical methods. In conclusion, this catalogue provides a tool for researchers to facilitate choosing the most valid PROM for studies in sports research. Furthermore, it shows for popular PROMs where further validation is needed, and for fields in musculoskeletal medicine where valid PROMs are lacking. It is suggested that a targeted effort is made to develop valid PROMs for major conditions in musculoskeletal research. The current method is easier to practice compared with assessment after COSMIN guidelines.

KEYWORDS

catalogue of PROMs, construct validity, content validity, patient-reported outcome measures, PROM, sports medicine, sports traumatology, validation

1 | INTRODUCTION

In clinical surveys, it is best to use a PROM that has been developed for the patients in question (eg patients with a traumatic meniscal lesion), and which has been found to be valid for these patients. Most articles on development and validation conclude that the particular PROM is a valid tool, and it is tempting just to refer to such an article when explaining the choice of PROM for a study. However, there are huge differences in the quality of development and validation

Christian F. Hansen and Jonas Jensen shared first authorship.

Case

A French research group is planning a randomized controlled trial (RCT) of the clinical effect of arthroscopic treatment compared with slow-heavy resistance training for patients with patellar tendinopathy. Which PROM should the group use?

The group searches PubMed to find what others have used. A search string containing the terms "patellar, tendinopathy, and RCT" reveals four relevant studies, and another string with "patellar, tendinopathy, and randomized" identifies additional studies. The group finds that among ten randomized studies, eight have used VISA-P as an outcome,[1-8] four used a VAS pain scale,[1,4,5,9] one used pain during activity,[6] and one used pain upon pressure pain thresholds.[10]

The group also finds that the development article for VISA-P[11] has concluded that VISA-P "is a reliable index of the severity of jumper's knee." Furthermore, an article concerning the validity of a local French version of VISA-P has concluded that the PROM "is understandable, valid, and suitable for French-speaking patients with patellar tendinopathy".[12]

The group chooses to use VISA-P for their study.

Comment: Even though VISA-P has been used in many clinical studies, and all validation studies conclude that it is a valid and reliable clinical measure for patients with patellar tendinopathy, VISA-P has never been confirmed to possess high content validity for the targeted patient population, nor has the construct validity ever been assessed using robust statistical methods. It is therefore unclear what it measures and how it functions.

studies, and, to most researchers, it is difficult to evaluate the usefulness of a PROM for a particular study in detail. It is often tempting to choose the same PROM others have used in similar studies. This has for instance been the case for many studies regarding the outcome of hip arthroscopy, even though the most commonly used condition-specific PROM, the modified Harris Hip Score (mHHS), was developed without involvement of patients and thus has not been subjected to content validation for the targeted patient group.[13]

The COnsensus-based Standards for the selection of health status Measurement Instruments (COSMIN) group recommends that development and validation of PROMs, which are candidates to be used as outcome measures in a particular study, are carefully assessed before a choice is made.[14] For many conditions, there are often several PROMs to choose from. Most literature reporting the development and validation of PROMs tends to be loyal to the PROM being described (ie the authors typically report that the PROM is valid and reliable). However, true assessment of validity is only possible by thorough review of the literature, which is time-consuming. Rarely do these papers report on the actual validation procedures used in the original development study.

The aim of this study was to produce a catalogue of the most widely used PROMs in sports medicine/traumatology research and conduct a thorough quality assessment of the development and validation of each PROM and each published local adaptation, including an indication of which patients each PROM is suited for. From this catalogue, it should be possible to identify the most appropriate PROM for studies involving sports medicine and sports traumatology. Furthermore, the aim was to continuously update this catalogue electronically.

2 | MATERIALS AND METHODS

To identify the most commonly used PROMs in relation to sports, a search was performed in PubMed September 22, 2017, with the search terms "patient reported outcome measures sports" without limits on time or journals. Hits were organized chronologically by publication date, starting with the latest articles, and abstracts were downloaded until a predefined number of 500 articles had been selected. The search was repeated on January 20, 2019, which revealed 270 extra articles, and on November 24, 2019, with an extra 145 articles.

The 915 abstracts and if necessary the full articles were manually studied. Articles describing conditions that are not relevant to sports (eg articles regarding hip osteoarthritis and arthroplasty), and articles that did not have at least one PROM (identified by name) as an outcome, were excluded. A total of 439 articles were included with the oldest article published July 29, 2011. A total of 194 different PROMs were identified, and the number of times each PROM had been included in the 439 articles was counted. From the identified PROMs, 42 PROMs that had been used three times or more were selected for the evaluation of their development and validation. Additionally, 13 PROMs that had been used only once or twice, but were the only PROMs for a particular condition, were also included in the selection. Furthermore, six PROMs that had not appeared in our search, although had been used in randomized controlled studies in sports research,[15,16] were also selected. In total, 61 condition-specific PROMs were selected for evaluation. Global PROMs and PROMs for mental/psychiatric conditions were not subjected to quality assessment.

The quality assessment of the development and validation of the selected PROMs was based on a review of

these processes. For each of the included PROMs, the articles that described how the PROM had been developed and originally validated were identified. In many cases, this was referred to in the articles that used the PROMs, but this was cross-checked by search in PubMed and SCOPUS. If it was not possible to identify any article, the person or organization that was associated with this PROM was contacted and the information that was necessary for assessments was requested. As PROMs are often validated multiple times, and for different patient groups, searches were conducted for subsequent validations of the original PROM and for translated/adapted versions in other languages with the search strategies "[PROM name] AND development", "[PROM name] AND validation", "[PROM name] AND translation" and "[PROM name] AND adaptation" in PubMed and SCOPUS. All the identified validations for each PROM were included if they were in English or Scandinavian language.

The information that was extracted from each article is listed in Table 1.

Quality assessment of the development and validation described in each article was performed independently by two authors (CFH and JJ) and discussed in plenary with VS and MRK.

2.1 | Quality assessment of the development

To operationalize the process of quality assessment of PROMs, a rating system was created for assessment of the level of quality from 1 to 5, where 1 represented the minimum level and 5 the optimal. Table 1 shows all the different components used to evaluate the quality of each PROM, and whether the quality indicators were present in the development and validation of the PROM (each component is indicated with a bullet in Table 1). For each quality indicator, a positive score was added to the overall assessments for development and validation. The quality indicator rating system was based on the generally accepted criteria for achieving high face and content validity.[14,17]

- The lowest score of one asterisk * would be given to a PROM based on a list of items developed without involvement of experts and patients and with no items from existing PROMs or an item bank.
- A PROM with items developed by a group of experts on the basis of a bank of items from existing PROMs of relevance would score two asterisks **, as the items would have face validity and be based on items that had been validated in existing PROMs.
- If the items had been discussed once in a group of patients with the health condition in focus, the score would be increased by one asterisk *.

TABLE 1 Information that was extracted from each article describing development and/or validation of a PROM

Name of the PROM
Year published and name of journal
Link to reference
Country where the study was performed
Language of the version in the article
Number of domains
Which condition was the PROM developed for (aim in development article) • Is there face validity in development • Is there content validity in development
Health condition of the patients who participated in development (number of patients) • Were items from existing PROMs used in development (name of PROM/PROMs) • Were items developed by help of patients in groups • Were items discussed until data saturation • Were items tested through interviews
Validation year and journal
Link to reference
Who validated (the same group who developed/an independent group)?
Language of the validated version of the PROM
Disease of the patients in validation (name of condition(s) and number in each group)
Were patients in different phases of disease (pre- and postoperatively) included (name(s) of phase(s) and number of patients)? • Were persons without the condition the PROM is to be used for included in the validation process? • Was a modern test theory method used for validation (Rasch model, other IRT models, confirmatory factor analyses) • Was the outcome compared to outcome with existing PROM(s)? • Were other validation methods used? • Was validation done on an aggregated score or on each domain separately? • Was test-retest reliability calculated? • Was Cronbach's α calculated? • Was known groups validity assessed? • Was differential item functioning assessed? • Was a minimal important difference calculated? • Was responsiveness assessed? • Were floor and ceiling effects investigated?
Author's conclusion
Our assessment of development and validation (including which health conditions the PROM was validated for)
Total assessment of the article with graded list of health conditions the PROM has been validated for

Bullet Indicates that this information was used in the quality assessment of the PROM.

- If items were discussed in focus groups until 'data saturation' (ie no further item themes emerged), the score would be increased by one asterisk * and

- If a provisional PROM had also been developed and tested through debriefing with the addition of debriefing interviews with individual patients with the health condition in focus, it would be increased by one asterisk *.

Therefore, the maximum score was five asterisks *****.

However, the criteria are to some degree independent and there is no ranking of them. This means that the scoring system should not be considered an invariant ordinal scale, but simply a pragmatic framework for classifying the quality of PROM development.

It should be emphasized that asking patients later whether the questions are relevant (either yes/no or by a grading scale), if they address their concerns, or whether there are unaddressed issues in a questionnaire (for instance, during construct validity assessment) cannot replace the development process via qualitative methodology in single or group interviews and therefore not adequately investigate the content validity.

2.2 | Quality assessment of validation

For the assessment of the quality of the validation, several factors were weighed:

- The number of patients used for validation: The physical conditions (eg shoulder instability, or meniscal injury) of the patients used for validation were registered. If patients were not included (for instance, if the article was only a translation of a PROM to a different language with no statistical validation) or the total number of patients was less than 20, the validation quality was scored with one asterisk * (lowest). Also, conditions with less than 20 patients (eg if 100 patients with five different conditions participated in the validation process, but one condition was represented by four patients only) were not mentioned in the conclusion of which conditions the PROM had been validated for.
- Psychometric methods for validation: Modern test theory (MTT) methods were regarded superior to classical test theory (CTT) methods.[18] All IRT methods (eg Rasch analysis) and confirmatory factor analysis (CFA) were rated as equal.
- Reliability: This can be assessed using different methods, such as Cronbach's α, test-retest, and Person Separation Index for MTT models. This was rated one asterisk *.
- Responsiveness: Validation of responsiveness and minimal important difference was rated one asterisk *.
- Aggregated sum scores: If tests in PROMs with several domains were only performed on the aggregated score (ie addition of all domain scores), this was regarded as a potential problem. Such scales often lack proven unidimensionality compared with tests on each separate domain score. This did not in itself change the assessment by a full asterisk, but it was considered in the total quality assessment of the validation.

The highest possible quality rating for validations that had not used IRT methods was three asterisks ***.

If IRT methods had been used, but nothing else, the validation was rated with four asterisks ****.

If IRT methods had been used and both reliability and responsiveness had been assessed, the validation was rated with five asterisks ***** (highest).

2.3 | Quality of the PROM validation (aggregated assessment)

2.3.1 | Methodological assessment

The aggregated score for each condition validated in each PROM study was basically the lowest of the assessments for development and validation. Therefore, a PROM which scored two asterisks in development could never end with an aggregated score of more than the two asterisks, even though it had been validated with MTT. A PROM which scored high in development (ie four or five asterisks), but had only been validated by CTT with a maximum score of three asterisks, would result in an aggregated score according to the assessment of validation. If, however, validation by MTT methods with a maximum score of five asterisks had been performed on a later occasion, the aggregated score could increase for the particular condition it was validated for.

2.3.2 | Quality of the instrument (overall aggregated assessment)

This assessment of the quality of the PROM, in addition to the methodological assessments, also took the results of validation studies into account, meaning: Did the results from a validation study that used an MTT model indicate that this is a PROM of high or low quality? For PROMs that have been validated through studies with an aggregated rating score of one, two, or three asterisks, it is not possible to establish how good the PROM actually is, because even though a PROM can be found valid by CTT, subsequent validation by MTT can reveal inadequate construct validity. Therefore, it was impossible to assess the quality of PROMs for which construct validity had not been evaluated by MTT.

In cases where MTT validation showed inadequate construct validity, for instance if unidimensionality of domains was not confirmed, or there was a floor/ceiling effect, the overall assessment was made after discussions in the study group.

There were three categories in the section "Quality of the instrument": High, low, and unknown.

- High: PROMs with a score of four or five asterisks in the aggregated methodological assessment based on one or

more studies, and with a positive validation by an MTT method, were characterized as high quality for the specific condition(s) that had been validated. Methodological assessment of one, two, or three asterisks could never characterize a PROM as a "high-quality instrument" for the health conditions it had been validated for.

- Low: PROMs for which an assessment with an MTT method showed weaknesses in the construct validity (eg lack of unidimensionality) were discussed in the group of authors, and in most cases, the PROM was characterized as a low-quality instrument for the condition(s) that had been assessed in the particular validation study, independently of the aggregated methodological assessment score. In few cases, the validity problems were minor, and the PROM could be characterized as a high-quality instrument if the methodological assessment score was four or five asterisks.
- Unknown: PROMs with validation studies in which construct validity had not been assessed by MTT were categorized as unknown overall quality for the condition(s) that had been assessed in these particular validation studies (no notation in Tables S1-S27). PROMs which had been found adequate by MTT, but had only scored one, two, or three asterisks in the quality assessment of development, would also be characterized as unknown overall quality (or low in cases where MTT revealed inadequate construct validity), as they have no confirmed content validity.

3 | RESULTS

The 61 PROMs that were assessed in the study are listed in Table 2. This also indicates the distribution according to bodily region of the 194 PROMs that were identified in the search.

Tables S1-S9 list the articles that describe development of each PROM, the details of this procedure, and for which condition(s) each PROM had been developed. The tables present PROMs for conditions in the neck, shoulder, elbow, hip, thigh, knee, calf, and foot/ankle, respectively. Tables S10-S18 list the quality assessments, including for which condition(s) each PROM had been validated. Tables S19-S27 list the overall quality assessment of each PROM and local translations.

References for Tables S1-S27 are listed separately in connection with each table.

4 | DISCUSSION

This study assessed the development and validation of 61 PROMs used for conditions in sports medicine and sports traumatology, and the published local translations of these PROMs. The results can serve as a catalogue for researchers when a relevant and thoroughly developed and validated PROM is to be chosen for a specific study.

TABLE 2 The 61 PROMs for which the development and validation are analyzed in this article (WOMAC is listed twice, as it is used for hip and knee separately, so the list includes 62 PROMs). In the headlines for each anatomical region it is stated how many PROMs that were identified in the search and from which the PROMs for the study were chosen. The number of times each PROM had been used in the articles of the search is listed in parentheses. * indicates that the PROM had only been mentioned once or twice but was the only PROM for a particular condition

Neck (number of PROMs identified: 2)
*[1]Neck disability index (1)
Shoulder (number of PROMs identified: 29):
American Shoulder and Elbow Surgeons (ASES) for shoulder patients (24)
Constant-Murley Score (12)
Quick Disabilities of the Arm, Shoulder and Hand (QuickDASH) (11)
Single Assessment Numeric Evaluation (SANE), used for shoulder (10)
Disabilities of the Arm, Shoulder and Hand (DASH) (8)
Western Ontario Rotator Cuff Index (WORC) (8)
Oxford Shoulder Score (OSS) (6)
PROMIS-UE (Upper Extremity) (6)
Western Ontario Shoulder Instability Index (WOSI) (6)
Rowe score (5)
Simple shoulder test (4)
Shoulder Pain and Disability Index (SPADI) (4)
Oxford Instability Shoulder Score (OISS) (3)
*[2]Short Western Ontario Rotator Cuff Index (short-WORC) (1)
Elbow (number of PROMs identified: 5):
Patient-Rated Tennis Elbow Evaluation (PRTEE) (4)
*[3]American Shoulder and Elbow Surgeons Elbow Questionnaire (pASES-e) (2)
Hand:
**Patient-Rated Wrist Evaluation (PRWE)
Hip (number of PROMs identified: 24):
Modified Harris Hip Score (75)
Hip Outcome Score (HOS) (55)
Non-arthritic Hip Score (NAHS) (30)
International Hip Outcome Tool-33 (iHOT-33) (22)
The Copenhagen Hip and Groin Outcome Score (HAGOS) (19)
Hip Disability and Osteoarthritis Outcome Score (HOOS) (18)
International Hip Outcome Tool-12 (iHOT-12) (13)
Hip Sports Activity Scale (HSAS) (9)
Oxford Hip Score (8)
Lower Extremity Function Score (LEFS) (8)
Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)—used for hip (6)
*[4]VISA-G (1)

(Continues)

TABLE 2 (Continued)

Thigh (number of PROMs identified: 3)
*5Functional Assessment Scale for Acute Hamstrings Injury (1)
*5Hamstring Outcome Score (1)
Knee (number of PROMs identified: 57):
International Knee Documentation Committee (IKDC) Questionnaire (103)
Knee Injury and Osteoarthritis Outcome Score (KOOS) (102)
Lysholm Score (71)
Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)—used for knee (24)
Knee Outcome Survey (KOS) (8)
Anterior Knee Pain Scale (Kujala Score) (7)
Modified Cincinnati Knee Rating Scale (6)
Hospital for Special Surgery Score (5)
Knee Society Score (4)
Victorian Institute of Sports Assessment-Patella (VISA-P) (4)
Knee Selfefficiacy Scale (4)
Knee Numeric-Entity Evaluation Score-ACL (KNEES-ACL) (4)
Pedi-International Knee Documentation Committee (Pedi-IKDC) Questionnaire (4)
*6Knee Injury and Osteoarthritis Outcome Score 4 (KOOS4) (2)
*7Forgotten Joint Score (FJS-12) (1)
*8KOOS-Child (1)
**Arthritis Impact Measurement Scales-2 (AIMS-2) (1)
Calf (number of PROMs identified: 2)
*9Lower Limb Functional Index (2)
*10Medial Tibial Stress Syndrome Score (MTSS Score) (1)
Ankle/foot (number of PROMs identified: 25):
Achilles tendon Total Rupture Score (ATRS) (15)
Foot and Ankle Outcome Score (FAOS) (14)
Foot and Ankle Ability Measure (FAAM) (11)
Victorian Institute of Sports Assessment-Achilles (VISA-A) (8)
Karlsson score (3)
*11Foot Function Index (FFI) (2)
*12American Orthopaedic Foot & Ankle Society Hindfoot Score (AOFAS-HAS) (2)
**American Academy of Orthopaedic Surgeons Foot and Ankle Outcomes Questionnaire (AAOS-FAOQ)
**Ankle Osteoarthritis Scale
**Kerr-Atkins score
**Leppilahti score
Global PROMs (number of PROMs identified: 42): Not evaluated.
Mental PROMs (number of PROMs identified: 5): Not evaluated.

*1 The only PROM for neck conditions.

*2 This is the short version of a commonly used PROM.

*3 The only general elbow PROM.

*4 The only PROM for greater trochanteric pain syndrome.

(Continues)

TABLE 2 (Continued)

*5 The only PROMs for hamstrings injuries.

*6 This PROM is a modification of a commonly used PROM.

*7 This PROM uses a different approach in items compared to other PROMs.

*8 The only European PROM for children with knee problems.

*9 The only PROM for general function of the lower leg.

*10 The only PROM for medial tibial stress syndrome.

*11 This PROM is for the general foot condition.

*12 This PROM is specific for the hindfoot.

** Indicates that the PROM had not appeared in our search, but was an outcome measure in the RCTs, analyzed elsewhere.[15,16]

The overall result is that the majority of these measurement instruments are of inferior validity, with few exceptions. The most prominent reason for this is that many PROMs have not been developed using methods that ensure content validity. It is generally agreed that a crucial prerequisite, and the most important quality of a PROM, is that it has high content validity.[14] Any deficits in this cannot be repaired by subsequent mathematical processes. Another major reason for inferior validity of PROMs is a lack of construct validation using adequate methods. However, if a PROM has good content validity, construct validation tests can be performed subsequently.

There are several review articles, web pages, and electronic databases, which describe assessment of various PROMs for sports medicine and sports traumatology. However, these sources have often referred directly to the conclusions from the development and validation articles, and have not provided a systematic approach in the quality assessment. Many articles describing development and validation of PROMs use suboptimal methods, and yet the vast majority conclude that the PROMs in question are valid tools. This contrasts our quality assessments, which show problems in both the development and validation (including responsiveness and reliability) of most musculoskeletal PROMs. Very few PROMs are rated as a high-quality instrument in our aggregated assessments.

This calls for a change in attitude for what should be expected of a PROM instrument. First, researchers should be critical in their choice of PROMs for a planned study. Second, journals publishing protocols should be critical of the validation of the measurement instruments in their review process. And third, journals should either not publish studies that have used PROMs of doubtful validity or should expect that any suspicion of suboptimal validity is declared and that the effect in the interpretation of the results is discussed.

This study was initiated in February 2017. In December 2017, COSMIN produced an e-published checklist for quality assessment of PROMs. The quality assessment principles in the current study are basically the same as recommended by the COSMIN group, but not as highly detailed. Both systems rank PROMs in few categories. Assessment by the COSMIN checklist can be very time-consuming for PROMs with a high

quality in development and validation, as the guidelines request many details, of which not necessarily all are readily available. It is a basic principle in our framework, as well as in the guidelines from COSMIN,[14] that the quality of a PROM is primarily dependent on its development. High-quality development ensures content validity, whereas low-quality development processes are difficult to correct through subsequent validation procedures. In a separate study, five PROMs, that had also been assessed in the current study, were assessed according to the COSMIN guidelines,[19] and from this, the two processes and the results hereof can be compared.

In our analyses, we have regarded exploratory factor analysis as inferior to MTT methods. We regarded confirmatory factor analysis as equivalent to IRT methods. This choice can be debated, but it is consistent with the COSMIN guidelines.[14]

The most relevant aspect when a PROM is chosen for a specific study is that development has followed optimal procedures and that the patient group in question has been involved. If validation (eg assessment of the construct validity) is not optimal for a PROM that was appropriately developed, it is possible to perform a validation (with MTT) on prospectively collected data, and we suggest that this is done in all cases where the PROM has not been validated by MTT methods for the actual patient group. However, such an assessment carries the risk that that the PROM turns out not to possess construct validity, and in such cases, this undermines the trustworthiness of the study results.

It is a paradox that the better the methods that are used to assess construct validity, the higher the probability that these methods can reveal a lack of construct validity. This means that PROMs that are assessed by suboptimal methods (ie CTT) cannot automatically be assumed to possess construct validity.

Short forms of existing PROMs can be developed by involving of patients or by data-driven methods (ie through mathematical analysis of data). If the PROM was developed with optimal patient involvement and therefore had coverage and relevance, and if it was validated with MTT methods, then items that did not add information would have been removed. As all items are not necessarily of equal importance, it is often possible to reduce the number of items or even domains by removing the least important items or domains. The risk of losing coverage is diminished if this is done with patient involvement. However, if it is done solely in a data-driven way, then there is an increased risk that items that contribute meaningful information about the condition might be removed. Also, the construction of short forms is highly dependent on the actual patient population, including factors like gender, age and specific health condition. The use of short forms is therefore extra critical in relation to the target patient population. This is also the reason why there exist several short forms with inclusion of different items from the original PROM (eg the Lower Extremity Functional Scale (LEFS) (Table S23)). In exploring whether a short form is useful for a specific study, the patient population that is the basis for this particular short form should be thoroughly analyzed for comparability with the patient population in question.

It seems strange that so few PROMs have been validated using robust statistical/mathematical methods that have been specifically developed for this purpose—MTT. It can be suspected that in some cases, MTT methods have in fact been used for validation but have shown a less than positive result and therefore the results have not been published.

BOX 1 How to use this catalogue - a quick guide

Step 1: Choose anatomical location (eg knee)

Step 2: Identify PROMs meant for the relevant condition in the development table (eg ACL injury in Table S7)

Step 3: Identify PROMs in your language (eg KNEES-ACL, IKDC, KOOS, Lysholm)

Step 4: Identify PROMs that have construct validity for the relevant condition (eg ACL injury in Table S16). If there is a PROM with studies rated "High" for the relevant condition, it should be preferred over PROMs with no quality rating or the rating "Low." If there is no PROM with studies of "High" quality, PROMs with higher rated studies might be preferable, but candidates should be assessed by reading the development and validation articles (which are referred in the tables)

Step 5: Check for updated validations, for instance, in PubMed

5 | CONCLUSION

From this catalogue, it is possible to identify the most valid PROMs for use as outcome measures for specific studies in sports medicine and sports traumatology. Box 1 is a quick guide as to use the catalogue. This catalogue is planned to be accessible electronically and be regularly updated.

6 | PERSPECTIVE

As a large proportion of PROMs in their development and documented validation turned out to be of low quality, and must be regarded as suboptimal measurement tools, it is important that a targeted effort is made to develop valid PROMs for major conditions.

CONFLICT OF INTEREST

All authors declare that they have no conflicts of interest in relation to this manuscript.

DATA AVAILABILITY STATEMENT

All data are published in the article and the supplementary material.

ORCID

Christian F. Hansen https://orcid.org/0000-0003-4769-4014
Volkert Siersma https://orcid.org/0000-0003-1941-2681
John Brodersen https://orcid.org/0000-0001-9369-3376
Jonathan D. Comins https://orcid.org/0000-0002-8284-114X
Michael R. Krogsgaard https://orcid.org/0000-0002-9976-4865

REFERENCES

1. Barker-Davies RM, Nicol A, McCurdie I, et al. Study protocol: a double blind randomised control trial of high volume image guided injections in Achilles and patellar tendinopathy in a young active population. *BMC Musculoskelet Disord*. 2017;18:204.
2. Gatz M, Schrading S, Dirrichs T, et al. Topical polidocanol application in combination with static stretching in tendinopathies: a prospective pilot study. *Muscles Ligaments Tendons J*. 2017;7:88-97.
3. Zwerver J, Verhagen E, Hartgens F, van den Akker-Scheek I, Diercks RL The TOPGAME-study: effectiveness of extracorporeal shockwave therapy in jumping athletes with patellar tendinopathy. design of a randomised controlled trial. *BMC Musculoskelet Disord*. 2010;11:28.
4. de Jesus JF, de Albuquerque TAB, Shimba LG, Bryk FF, Cook J, Pinfildi CE High-energy dose of therapeutic ultrasound in the treatment of patellar tendinopathy: protocol of a randomized placebo-controlled clinical trial. *BMC Musculoskelet Disord*. 2019;20:624.
5. Lee WC, Ng GY, Zhang ZJ, Malliaras P, Masci L, Fu SN Changes on tendon stiffness and clinical outcomes in athletes are associated with patellar tendinopathy after eccentric exercise. *Clin J Sport Med*. 2020;30:25-32.
6. Scott A, LaPrade RF, Harmon KG, et al. Platelet-rich plasma for patellar tendinopathy: a randomized controlled trial of leukocyte-rich PRP or leukocyte-poor PRP versus saline. *Am J Sports Med*. 2019;47:1654-1661.
7. Lee WC, Ng GY, Zhang ZJ, Malliaras P, Masci L, Fu SN Changes on tendon stiffness and clinical outcomes in athletes are associated with patellar tendinopathy after eccentric exercise. *Clin J Sport Med*. 2020;30:25-32.
8. Abat F, Sánchez-Sánchez JL, Martín-Nogueras AM, et al. Randomized controlled trial comparing the effectiveness of the ultrasound-guided galvanic electrolysis technique (USGET) versus conventional electro-physiotherapeutic treatment on patellar tendinopathy. *J Exp Orthop*. 2016;3:34.
9. Zhao R, Liu K, Li P Clinical observation of moxibustion plus exercise prescription for patella tendinopathy of athlete. *Zhongguo Zhen Jiu*. 2016;36:927-930.
10. Holden S, Lyng K, Graven-Nielsen T, et al. Isometric exercise and pain in patellar tendinopathy: a randomized crossover trial. *J Sci Med Sport*. 2019;10:S1440-2440(19)30855-2.
11. Visentini PJ, Khan KM, Cook JL, Kiss ZS, Harcourt PR, Wark JD The VISA score: an index of severity of symptoms in patients with jumper's knee (patellar tendinosis). Victorian Institute of Sport Tendon Study Group. *J Sci Med Sport*. 1998;1:22-28.
12. Kaux J-F, Delvaux F, Oppong-Kyei J, et al. Cross-cultural adaptation and validation of the Victorian Institute of Sport assessment-patella questionnaire for french-speaking patients with patellar tendinopathy. *J Orthop Sports Phys Ther*. 2016;46:384-390.
13. Kemp JL, Collins NJ, Roos EM, Crossley KM Psychometric properties of patient-reported outcome measures for hip arthroscopic surgery. *Am J Sports Med*. 2013;41:2065-2073.
14. Mokkink LB, de Vet HCW, Prinsen CAC, et al. COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018;27:1171-1179.
15. Hansen CF, Jensen J, Siersma V, Brodersen J, 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;31:972-981.
16. Jensen J, Hansen CF, Brodersen J, Comins JD, Krogsgaard MR Are PROMs used adequately in sports research? An analysis of 54 randomized controlled trials with PROMs as endpoint. *Scand J Sci Med Sports*. 2021;31:982-990.
17. Comins JD, Brodersen J, Siersma V, Jensen J, Hansen CF, Krogsgaard MR. How to develop a condition-specific PROM. *Scand J Med Sci Sports*. 2020. https://doi.org/10.1111/sms.13868
18. Christensen KB, Comins JD, Krogsgaard MR, et al. Psychometric validation of PROM instruments. *Scand J Med Sci Sports*. 2020. https://doi.org/10.1111/sms.13908
19. Hansen CF, Jensen J, Odgaard A, et al. Validity of five commonly used PROMs in sports research evaluated by the COSMIN Risk of Bias Checklist - KNEES-ACL, HAGOS, KOOS, IKDC, and modified Harris' Hip Score. *Knee Surg Sports Traumatol Arthrosc* 2021.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: 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;31:991–998. https://doi.org/10.1111/sms.13923