

Patient-reported outcome measures for hip-related pain: a review of the available evidence and a Consensus Statement from the International Hip-related Pain Research Network, Zurich 2018

Franco M. Impellizzeri¹, Denise M Jones², Damian Griffin³, Marcie Harris Hayes⁴, Kristian Thorborg⁵, Kay M Crossley⁶, Michael P Reiman⁷, Mark Scholes⁸, Eva Ageberg⁹, Rintje Agricola¹⁰, Mario Bizzini¹¹, Nancy Bloom¹², Nicola C Casartelli¹³, Laura E Diamond¹⁴, H Paul Dijkstra¹⁵, Stephanie Di Stasi¹⁶, Michael K Drew¹⁷, Daniel J Friedman¹⁸, Matthew Freke¹⁹, Boris Gojanovic²⁰, Joshua J Heerey²¹, Per Hölmich²², Michael A Hunt²³, Lasse Ishøi²⁴, Ara Kassarian²⁵, Matthew G King²⁶, Peter R Lawrenson²⁷, Michael Leunig²⁸, Cara L Lewis²⁹, Kristian Marstrand Warholm³⁰, Susan Mayes³¹, Håvard Moksnes³², Andrea B Mosler³³, May-Arna Risberg³⁴, Adam I Semciw³⁵, Andreas Serner³⁶, Pim van Klij³⁷, Tobias Wörner³⁸, Joanne L Kemp³⁹

1,

Faculty of Health,
University of Technology Sydney
Sydney, Australia

2,

La Trobe Sport and Exercise Medicine Research Centre
La Trobe University, Melbourne, Australia

3,

Professor of Trauma and Orthopaedic Surgery
Warwick Medical School, University of Warwick,
Warwick, UK

4,

Physical Therapy and Orthopaedic Surgery
Washington University School of Medicine

Saint Louis, USA

5,

IOC Research Center Copenhagen

Sports Orthopedic Research Center - Copenhagen (SORC-C)

Arthroscopic Center, Department of Orthopedic Surgery

Copenhagen University Hospital, Amager-Hvidovre

Denmark

6,

La Trobe Sport and Exercise Medicine Research Centre

La Trobe University, Melbourne, Australia

7,

Division of Physical Therapy

Duke University Medical Center

Durham, NC. 27705

8,

La Trobe Sport and Exercise Medicine Research Centre

La Trobe University, Melbourne, Australia

9,

Department of Health Sciences

Lund University

Sweden

10,

Department of Orthopaedic Surgery, Erasmus MC, University Medical Centre Rotterdam,

Rotterdam, the Netherlands

11,

Schulthess Clinic

Human Performance Lab

Zürich Switzerland

12,

Physical Therapy and Orthopedic Surgery

Washington University School of Medicine

St. Louis MO. USA

13,

Human Performance Lab, Schulthess Clinic, Zurich, Switzerland

Laboratory of Exercise and Health, ETH Zurich, Zurich, Switzerland

14,

Menzies Health Institute Queensland, School of Allied Health Sciences,

Griffith University, Queensland, Australia

15,

Weill Cornell Medicine Qatar

Doha, Qatar

16,

Division of Physical Therapy, School of Health and Rehabilitation Sciences

Sports Medicine Research Institute, The Ohio State University Wexner Medical Center

17,

Athlete Availability, Australian Institute of Sport, Australia

University of Canberra Research into Sport and Exercise, University of Canberra, Australia

Australian Centre for Research into Injury in Sport and its Prevention (ACRISP), Australia

18,

School of Medicine, Monash University,
Melbourne, Victoria, Australia

19,

School of Health & Rehabilitation Sciences, The University of Queensland,
Brisbane, Queensland 4072 Australia

20,

Health and Performance Medical Director, Swiss Olympic Medical Center, Hôpital de La Tour,
Meyrin, Suisse

21,

La Trobe Sport and Exercise Medicine Research Centre
La Trobe University, Melbourne, Australia

22,

IOC Research Center Copenhagen
Sports Orthopedic Research Center - Copenhagen (SORC-C)
Arthroscopic Center, Department of Orthopedic Surgery
Copenhagen University Hospital, Amager-Hvidovre
Denmark

23,

Department of Physical Therapy,
University of British Columbia
Vancouver, Canada

24,

IOC Research Center Copenhagen
Sports Orthopedic Research Center - Copenhagen (SORC-C)

Arthroscopic Center, Department of Orthopedic Surgery
Copenhagen University Hospital, Amager-Hvidovre
Denmark

25,
Founder and Consultant Musculoskeletal Radiologist
Elite Sports Imaging, SL
Madrid, Spain

26,
La Trobe Sport and Exercise Medicine Research Centre
La Trobe University, Melbourne, Australia

27,
School of Health & Rehabilitation Sciences, The University of Queensland,
Brisbane, Queensland 4072 Australia

28,
Orthopaedic Department, Schulthess Clinic
Zürich Switzerland

29,
Department of Physical Therapy & Athletic Training
College of Health & Rehabilitation Sciences: Sargent College
Boston University,
Boston, USA

30,
Oslo University Hospital
Oslo, Norway

31,

The Australian Ballet, Southbank, Melbourne, Australia

La Trobe University, Melbourne, Australia

32,

Oslo Sports Trauma Research Center,

Oslo, Norway

33,

La Trobe Sport and Exercise Medicine Research Centre

La Trobe University, Melbourne, Australia

34,

Oslo University Hospital and

Oslo Sports Trauma Research Center,

Oslo, Norway

35,

La Trobe Sport and Exercise Medicine Research Centre

La Trobe University, Melbourne, Australia

36,

Aspetar Orthopaedic & Sports Medicine Hospital

Doha, Qatar

37,

Department of Orthopaedic Surgery, Erasmus MC, University Medical Centre Rotterdam,

Rotterdam, the Netherlands

38,

PhD student, M.Sc.

Department of Health Sciences

Lund University, SWEDEN

39,

La Trobe Sport and Exercise Medicine Research Centre

La Trobe University, Melbourne, Australia

Corresponding author:

Franco M. Impellizzeri

Faculty of Health, University of Technology Sydney

Driver Avenue, Moore Park NSW 2021

PO Box 123. Broadway NSW 2007 Australia

T. +61 (02) 9514 5204

M. +61 (0) 404 560 633

E-mail: franco.impellizzeri@uts.edu.au

ABSTRACT

Hip-related pain is an increasingly recognised complaint in young and middle-aged active adults. People experiencing hip-related disorders commonly report pain and reduced functional capacity, including difficulties in executing activities of daily living. Patient reported outcome measures (PROMs) are essential in order to accurately examine and compare the effects of different treatments on disability in those with hip pain. In November 2018, 38 researchers and clinicians working in the field of hip-related pain met in Zurich, Switzerland for the 1st International Hip-related Pain Research Network meeting. Prior to the meeting, evidence summaries were developed relating to four prioritised themes. This paper discusses the available evidence and consensus process from which recommendations were made regarding the appropriate use of PROMs to assess disability in young and middle aged active adults with hip-related pain.

The process for gaining consensus was developed according to five steps: 1) systematic reviews of systematic reviews; 2) preliminary discussion within the working group; 3) update of the more recent high-quality systematic review and examine the psychometric properties of PROMs according to established guidelines; 4) formulation of the recommendations considering the limitations of the PROMs derived from the examination of their quality; and 5) voting and consensus. Out of 102 articles retrieved, 6 systematic reviews were selected and assessed for quality according to AMSTAR 2. Two showed a moderate quality. We then updated the most recent review. The updated literature search resulted in 10 additional studies that were included in the qualitative synthesis. The recommendations based on evidence summary and PROMs limitations were presented at the consensus meeting.

The following statements and recommendations were made by the group: (i) the HAGOS and iHOT instruments (long and reduced versions) are the most appropriate PROMs to use in young and middle-aged active adults with hip-related pain; (ii) more research is needed into their utility in a non-surgical treatment context; (iii) generic quality of life measures such as EQ-5D and SF-36 may be

a useful addition; and (iv) due to none of the instruments showing acceptable quality of all psychometric properties, more methodological studies are needed to further evaluate the validity of the HAGOS and IHOT, and the other non-recommended PROMS .

INTRODUCTION

Hip-related pain is an increasingly recognised complaint in both young and middle-aged active adults and athletes¹⁻³. In these populations, hip disorders are associated with increased disability as defined by the International Classification of Functioning, Disability and Health (ICF) developed by the World Health Organisation (<https://www.who.int/classifications/icf/en/>)^{4,5}. According to this biopsychosocial model, disability involves dysfunctions at one or more of the following three levels: impairment, activity limitations and participation restriction⁶. Indeed, people suffering from hip-related disorders commonly experience pain, impairments of body function and structure, and difficulties when executing activities of daily living and sports⁷.

In order to examine and compare the effects of different treatments on disability, it is necessary to use patient reported outcome measures (PROMs). Currently, PROMs are considered a necessary aspect of medical treatment evaluation^{8,9}, and are used in national and international registries¹⁰. Furthermore, PROMs are frequently used and recommended to support clinical decision-making, health policies and reimbursement processes¹¹. This requires the systematic collection of PROMs in the clinical setting. For these purposes, PROMs need to be valid and possess adequate psychometric properties. Lack of validity or suboptimal measurement properties of the PROMs might bias (positively or negatively) the effects of randomised-controlled trials¹². The respondent and patient burden of the selected PROMs must also be considered for successful implementation in research and clinical practice⁹. Given the importance of using appropriate PROMs, internationally recognised guidelines such as the Consensus-based Standards for the selection of health Measurement Instruments (COSMIN) were developed (<https://www.cosmin.nl/>). The COSMIN initiative targets improving the quality of studies investigating measurement properties. By developing methodology and practical tools for assessing measurement properties, the COSMIN guidelines can be used by clinicians and researchers to select the most appropriate instruments.

The aim of this paper was to present the consensus reached at the 1st International Hip-related Pain Research Network (IHiPRN) Consensus Meeting (November 2018, Zurich, Switzerland) on the most appropriate PROMs to assess disability in young and middle aged active adults with hip-related pain in both research and clinical settings.

METHODS

Consensus process

The first step of the five step process for gaining consensus included a systematic review (SR) of the SRs to define the best PROMs based on available literature. After examination of the quality of the selected SRs, the working group decided to update the most recent high-quality review. We assessed the quality of the psychometric properties of the PROMs recommended by Thorborg et al.¹³ and those identified in our update of this SR. Based on the quality and limitations of the PROMs obtained from the update and quality assessment, recommendations were developed for voting and consensus.

Step 1. Systematic review of the systematic reviews

Eligibility, inclusion and exclusion criteria

Peer-reviewed SR examining the psychometric properties of PROMs for patients with hip-related pain that included the following: (population) patients with hip pain, (including hip osteoarthritis and FAI syndrome and groin pain), (measurement properties) all measurement properties in any clinical context (surgery, non-surgical, etc.), (instrument) patient reported outcome measures.

Search strategy

The literature search was conducted in MEDLINE, Cochrane Database of SRs, Database of Abstracts of Reviews of Effects, Web of Science (no language or date restrictions: all articles before 31st July 2018) according to the following strategy adapted for each database: #1 (hip) OR (groin) OR (inguinal AND hernia); #2 (outcome AND assessment*) OR (self AND assessment*) OR (questionnaire*) OR (patient AND reported AND outcome*) OR (self AND report*); #3 (psychometric AND property*) OR (validity) OR (clinimetrics); #4 (systematic AND review); #5 #1 AND #2 AND #3 AND #4

Selection, data extraction and assessments

Title, abstract and full text were screened and aim, population, context/setting, number of instruments, suggested instruments and main authors' conclusions were extracted (Appendix 1). The screening, selection, data extraction and assessments of the SRs were carried out by two reviewers (FMI, DJ) and with a third (JK) acting as referee to solve conflicts. While for study selection there was a substantial agreement, for quality assessment kappa coefficient was fair to moderate ($k < 0.40$). This was the consequence of difficulties in the interpretation of the new COSMIN guidelines. Therefore discussion for solving and addressing sources of conflicts was necessary. This harmonisation improved the agreement between reviewers ($k > 0.76$)

Quality assessment

The quality of SRs was assessed using A MeaSurement Tool to Assess systematic Reviews (AMSTAR2, <https://amstar.ca/Amstar-2.php>), adapted to the topic of the SR that included studies investigating the psychometric properties of questionnaires. Specifically, item 14 (“Did the review authors provide a satisfactory explanation for, and discussion of, any heterogeneity observed in the results of the review?”) was not considered applicable because the heterogeneity of the results in methodological studies is not checked quantitatively, and some heterogeneity in the results is expected since psychometric properties are population and context specific. Similarly, items 11, 12 and 15 were not considered applicable since no quantitative meta-synthesis has been performed in the reviews.

Results of systematic review of systematic review

After duplicate removals, 102 articles were screened from titles and abstracts. 14 full texts were selected.¹³⁻²⁶ Eight literature reviews were excluded and 6 included (see articles in Table 1).^{13 17 20 24-26} The flow diagram (Figure 1) was presented according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA). The quality of the SRs (Table 1 and Appendix 1) according to the AMSTAR 2 was often deemed ‘critically low’ mainly because very few reviews assessed and took into consideration the risk of bias and/or methodological quality of the studies included in the SRs. Only two reviews were rated as moderate quality: Tijssen et al.²⁵ and Thorborg et al.¹³ These two SRs examined the quality of the methodological studies included and the PROMs according to an older version of the COSMIN (www.cosmin.nl). Based on their quality assessment, Tijssen et al.²⁵ recommended the use of Nonarthritic Hip Score (NAHS) and Hip Outcome Score (HOS). The review by Thorborg et al. was an update of their previous SR published in 2011 where they also recommended the NAHS. However, in their update, Thorborg et al.¹³ excluded the NAHS and they recommended the HOS, the Hip and Groin Outcome Score (HAGOS) and the International Hip Outcome Tool (iHOT-12 and iHOT-33) since these were the PROMs with the smallest proportion of specific psychometric properties with a poor methodology score.

Figure 1.

Table 1.

Appendix 1.

Step 2. First round discussion among the working group participants

The results of the SR of the SRs were circulated among the members of the working group. We decided to update the review by Thorborg et al.¹³ that was deemed the SR with the higher quality, the most recent and specifically focused on the target population of this consensus.

Step 3. Update of the systematic review by Thorborg et al.

The eligibility, exclusion and inclusion criteria and search strategy were the same as used in the review by Thorborg et al.¹³, but with dates modified to include only studies from 2015 to July 31st 2018.

Results of the update

Out of 803 articles found, 20 full texts were assessed for eligibility²⁷⁻⁴⁶ and 10 studies were included in the qualitative synthesis^{27 28 32 36 37 39-41 45 46}. The flow chart of the literature search for the updated SR is presented in Figure 2. We excluded the studies by Brans et al.²⁹ and Stevens et al.⁴³ since the mean age of the samples was higher (51-52 years) than 50 years old set as upper limit by our inclusion criteria. This replicated the inclusion criteria of Thorborg et al.¹³ to ensure consistency. We conducted a sensitivity analysis to examine whether these two papers could have influenced the overall rating. The assessment suggested that these two papers would not substantially change the final recommendations and were consistent with those included in the summary assessment (Appendix 2).

Figure 2.

Appendix 2.

Quality of the studies

For all papers included in the updated SR, the quality of the studies and of the psychometric properties were evaluated using the most recent COSMIN manual (version 1.0 updated February 2018, https://cosmin.nl/wp-content/uploads/COSMIN-syst-review-for-PROMs-manual_version-1_feb-2018.pdf). As per the SR of SRs, the same three reviewers were involved in the study selection, quality assessment and data extraction of all included articles.

The 10 selected studies examined five PROMs. Two PROMs (Core Outcome Measure Index and Oxford Hip Score) that were developed for other conditions (back pain and patients undergoing total hip arthroplasty respectively) were evaluated for their performance in FAI population^{36 37}. These

were assessed for their quality but given they address very few psychometric properties and their content validity in hip-pain patients was not evaluated, they were excluded from further analysis (i.e. quality assessment of the measurement properties). The remaining three PROMs were among the four recommended by Thorborg et al.¹³ To provide a summary of the quality of the evidence, studies in the updated SR were combined with studies examining the same PROMs (iHOT-33, iHOT-12, and HOS) reported in the review by Thorborg et al.¹³ For consistency, since the updated SR used a different version of the COSMIN manual, the assessments undertaken by Thorborg et al were redone using the last version of the COSMIN manual. (<https://cosmin.nl/wp-content/uploads/COSMIN-methodology-for-content-validity-user-manual-v1.pdf>).

Table 2.

Table 3.

Critical issues in rating the quality of the studies

There were critical issues relating to the rating of the quality of the studies for structural validity, internal consistency and cross-cultural validity. These are presented in Appendix 3.

Appendix 3.

Quality of the measurement properties

The quality of the measurement properties were rated for the instruments recommended by Thorborg et al. (HAGOS, HOS, iHOT12 and-33). The psychometric properties in the studies reported in the previous SR by Thorborg et al.¹³ (Table 4) were also reassessed.

The overall rating for structural validity reflected the lack of consistency in structure evaluation as mentioned in the previous section. In addition, no studies reported any fit indices, which is a requirement for assigning a positive rating using the COSMIN criteria. The measurement error was consistently higher than the minimal important change thus resulting in a negative rating. Finally, the updated COSMIN now allows the reviewers to develop the hypotheses (for construct validity and responsiveness), even if these are not explicitly stated by the authors. This resulted in more positive ratings, but this approach makes the assessment quite reviewer dependent and somewhat arbitrary.

Table 4.

Content validity

We evaluated the content validity using the new purposely developed COSMIN manual (<https://cosmin.nl/wp-content/uploads/COSMIN-methodology-for-content-validity-user-manual-v1.pdf>). The COSMIN manual suggests three steps in assessing the content validity of the PROMS and the quality of the corresponding studies: 1) evaluate the quality of the PROM development, 2) evaluate the quality of content validity studies, and 3) evaluate the content validity of the PROM. The analysis of content validity was performed for the iHOT33 and HAGOS only. The HOS did not involve any patients in the development phase and therefore as the content validity was considered inadequate, it was excluded from further examination. The COSMIN suggests that a modified PROM should, in principle, be treated as a new instrument. However, COSMIN also states that if the PROM is a modified shortened version, the information can be taken from the original PROM. This is the reason why the iHOT-12 (shorter version of the iHOT-33) was included among the recommended PROMs despite the content validity of this shorter version not being addressed specifically.

1. Evaluate the quality of the PROM development

Based on the worst score approach, the overall rating for the quality of PROM design to ensure relevance was inadequate because inadequate was the first item addressing the description of the construct to be measured. Indeed, both HAGOS and iHOT-33 did not describe or provide any operational definitions of the constructs. This increases the difficulty in interpreting whether the items of the PROMs are relevant for the construct of interest. In addition, while the HAGOS referred to the ICF framework and the inclusion of body structure, function and participation, the iHOT-33 did not report any theoretical grounding. The HAGOS used the Hip dysfunction and Osteoarthritis Outcome Score (HOOS) as template and reported the constructs of symptoms, pain, physical activity, sport and quality of life. However, detail on the aspects of these constructs in HAGOS was not provided. For example, both HOOS and HAGOS purport to assess pain however the dimensions of pain (pain intensity or interference) are not described. Quality of life is another broad and multifaceted concept included in the HAGOS and HOOS and a clear description would be necessary, but is not reported. Examination of the items suggests that other dimensions of quality of life have been considered compare to those addressed by traditional generic quality of life questionnaires

such as the EuroQoL (EQ5D), Short Form Health Survey (SF-36 and SF-12) or World Health Organization Quality of Life Instruments (WHOQOL). Most items relative to the methodological approach were rated as doubtful as clear descriptions of the methods were lacking.

2. Evaluate the quality of content validity studies

The COSMIN manual suggests that studies with a translation of a PROM should include a pilot study following translation, to evaluate the comprehensibility of the translated PROM. All cross-cultural validation studies did not formally report pilot studies to examine comprehensibility. At best they mentioned that comprehensibility was addressed in groups of patients, but without reporting any methods or results. For this reason, these studies were not considered as content validity studies and hence were excluded.

3. Evaluate the content validity of the PROM

Content validity was assessed using only the PROM development study. The reviewers' ratings were quite positive mainly based on the assumption that, even if not reported, some issues were probably addressed. This evaluation was subjective and based on an arbitrary interpretation of the items and response options included in the PROMs. The main problem of PROM development studies was that too few details about the content validity process were reported, such as how interviews were conducted, recorded and coded (e.g. use of nVIVO), and the reference framework for data extraction and coding. Details of the content validity assessment according to the COSMIN manual are presented in the Appendix 4. The evidence synthesis of the content validity is reported in Table 5.

Table 5.

Appendix 4.

Step 4. Formulation of the recommendation, including background and process of the consensus meeting

Selection of expert group members

The IHiPRN leadership group (Joanne Kemp, Kay Crossley, Mario Bizzini, Andrea Mosler, Cara Lewis and Karim Khan) met in January 2017 to set criteria to identify potential expert group members. Experts were selected based on their previous publications, and being current active researchers in the field of hip-related pain in young and middle-aged adults. Researchers who were also clinicians in the field were viewed favourably. Potential expert group members were contacted via email asking them an expression of interest in taking part in the 1st IHiPRN Consensus Meeting in Zurich in November 2018. Potential expert group members were also asked to suggest other experts for invitation that the leadership group may not have identified.

Following this expression of interest, four key areas were identified as priorities for consensus. These four key areas were:

- I. Classification of hip pain (including use of (i) clinical tests; and (ii) imaging)
- II. Patient-reported outcome measures for hip pain (including hip- related measures, and maybe others including pain / coping / fear / utility measures)
- III. Standardised measurement of physical capacity in hip-related pain (including clinical measures, biomechanics, EMG, physical activity, functional performance, and return to sport)
- IV. Physiotherapist-led treatment of hip-related pain.

The leadership group then identified experts to lead each of the four working groups. These were MR and RA (group 1), ABM and CLL (group 2), FI and JLK (group 3), JLK and MB (group 4). This paper relates to working group 2. Members of the working groups were then determined following discussion between the leadership group and the working group leaders.

Expert group demographics

All consensus meeting participants were considered to be experts and at the time of meeting were actively researching in the field of hip-related pain in young and middle-aged active adults. Areas of expertise among the participants included physiotherapy, orthopaedic surgery, sport and exercise medicine, biomechanics, diagnostics, imaging and radiology, patient-reported outcome measures, and exercise science. In addition, many of the participants were also expert clinicians who regularly treat young and middle-aged active adults with hip-related pain.

Step 5: Consensus process

The evidence summaries and draft recommendations were e-mailed to the delegates, at least two weeks prior to the meeting in Zurich. At the meeting, each working group met to discuss recommendations and revisions were made based on the discussion. The evidence summary and revised recommendations were presented to the expert group, with opportunity for discussion. The recommendation was then revised and finalized. At the conclusion of the discussion, each delegate was asked to vote on the recommendation. The voting was conducted anonymously, using a scoring system used at previous consensus meetings.^{47 48} A 10-point Likert scale was used to score each recommendation, where 0 was considered to be “inappropriate” and 9 “appropriate”. As described previously^{47 48}, scores were pooled and the median (interquartile range (IQR)) for each recommendation was determined. Scores that were 0-3 were considered inappropriate, scores 4-6 considered uncertain, and scores 7-9 were considered appropriate. Consensus statements were then developed based on the level of evidence available combined with the pooled voting score for that statement.

RESULTS

The consensus meeting in Zurich, Switzerland on November 17 and 18 2018, was attended by 37 participants. In addition, six participants were not able to attend in person, one attended the meeting via videoconferencing. Thus 38 participants were involved in the consensus voting process. All delegates were considered to be experts and were actively researching in the field of hip-related pain in active adults. Areas of expertise within the delegates included physiotherapy, orthopaedic surgery, sport and exercise medicine, biomechanics, diagnostics, imaging and radiology, patient-reported outcome measures and exercise science. In addition, many of the delegates were also clinicians treating adults with hip-related pain.

The median scores (IQR) for the 4 statements were 9(8) points. The scores for each statement are presented in Figure 3 and the final for recommendations of consensus group are presented in table 6.

Table 6.

Figure 3.

DISCUSSION

Recommendation 1. The HAGOS and iHOT instruments (long and reduced versions) are the most appropriate PROMs to use in young and middle-aged active adults with hip-related pain.

Based on the updated literature review and the quality assessment of the psychometric properties, we partially confirmed the suggestions by Thorborg et al.¹³ who recommended the HAGOS, HOS, iHOT-12 and iHOT-33. We excluded the HOS because this instrument was developed without the involvement of patients, which is necessary for ensuring content validity. Unfortunately, no subsequent studies examined the content validity of the HOS.

The HAGOS and the iHOT instruments (with iHOT-12 considered as a short version of the iHOT-33) had sufficient quality (mostly with high evidence) for cross-cultural validity, reliability and construct validity. The structural validity rating of all recommended PROMS was indeterminate, because the structure of the subscales and not the whole instruments was examined. The internal consistency of the subscales was sufficient with high evidence for HAGOS and low to moderate evidence for iHOT. High quality studies, however, showed large measurement error for both HAGOS and iHOT, where the smallest detectable change was higher than the minimal clinical important change (when available). Therefore, the usefulness of HAGOS and iHOT to evaluate the response to treatment of individual patients over time seems to be limited.

Although we excluded the HOS from the recommended instruments, we acknowledge that the other psychometric properties of the HOS were comparable to the other instruments. Therefore, despite its exclusion, the HOS may be potentially appropriate for this population if the content validity is confirmed in the future.

Recommendation 2. HAGOS and iHOT were developed mainly in surgical context. More research is needed into their utility in a non-surgical treatment context.

The HAGOS and iHOT have only been investigated in a surgical context (patients assessed before and after surgical interventions) or in mixed populations (undergoing both surgical and non-surgical treatments) (see details on population and context in Appendix 1). The magnitude of the effects following surgical interventions is not necessarily comparable with non-surgical treatment, which can impact the acceptability of measurement error and instrument responsiveness. Since the acceptability of reproducibility level (instrument noise) depends on the context and the magnitude of changes determined by the interventions (signal), we recommended the HAGOS and iHOT-33 primarily as outcome measures in a surgical setting (which is the main context in which they were investigated), while in non-surgical treatment the aforementioned limitations should be taken into consideration.

Recommendation 3. EQ-5D and SF-36 are generic quality of life measures that can supplement the hip-related measures, HAGOS and iHOT.

The use of generic questionnaires, in addition to condition specific PROMs, is commonly suggested to give a more complete picture of the patient health status.⁴⁹ These instruments were developed to be used with a generic population. There are several generic instruments available and the selection of a generic questionnaire for use in a particular clinical population should be based on theoretical considerations (e.g. what aspects of quality of life are of interest or whether a utility questionnaire is needed). For these reasons it is difficult to recommend a specific generic instrument. However, in absence of a gold standard instrument, it is common to use generic questionnaires to examine the construct validity (convergent evidence and hypotheses generation). For example, EuroQoL(EQ)-5D⁵⁰⁵¹ and SF36⁵²⁵³ are the generic instruments most commonly used as reference for the HAGOS and iHOT. These two instruments can be suitable generic questionnaires to use in addition to HAGOS and iHOT considering that they also provide health utility measures⁵⁴ and comparative values for hip-related pain population are available.

Recommendation 4. Future research should include further analysis of content and structural validity, and the relationship between individual measurement error and the minimal clinically important change for the recommended PROMs.

The examination of study quality and measurement properties highlighted inadequate structural validity, meaning that the structural validity of PROMs could not be determined despite us recommending their use. The structure of HAGOS⁵⁵ was developed using the Hip Disability and Osteoarthritis Outcome Score (HOOS) as a template⁵⁶, and not with a confirmatory analysis, but the HOOS structure was also not examined, but based on the structure of the Knee Injury and Osteoarthritis Outcome Score (KOOS)⁵⁷. Since the KOOS structure was not examined, a SR on the KOOS psychometric properties scored the structural validity as “poor” (according to the COSMIN)⁵⁸. Similarly, the structure of the iHOT was not properly examined or confirmed. Lack of structural validity examination is an important weakness, especially for instruments providing a single score such as the iHOT, as this limits interpretation of the total score. The operational definitions and theoretical framework of the construct reflected by the subscales was also not specified for the HAGOS and iHOT. These limitations are reflected in the content validity score. Despite being rated as sufficient by the reviewers, the content validity was mostly deemed to be inconsistent or indeterminate due to the lack of methodological information. Therefore, future studies should examine the structural validity, clarify the constructs measured and analyse the content validity of the HAGOS and iHOT. Finally, the measurement error was higher than the minimal clinically important change, thus questioning the use of these PROMs at individual level (e.g. in clinical practice), particularly for the iHOT. While the measurement error may be sufficient to detect change over time at a group level (e.g. research studies), further studies are needed to examine the minimal clinical change and its relationship with measurement error at individual level, especially for the iHOT.

Limitations

The expert group were from Europe, North America and Australia/New Zealand, limiting the cultural diversity of the group. Also, there were more men than women in the expert group and no patients were involved. Future meetings should try to improve all types of diversity and involve all stakeholders. While the use of the COSMIN manuals provided reference guidelines to assess the quality of the studies and the measurement properties of the PROMs, the interpretation of the items and hence the scoring is reviewer dependent. However, COSMIN acts as a guideline (as also stated in the manual) and allows for a certain degree of interpretation. This might influence our quality assessment results and the corresponding recommendations. Nevertheless, we used systematic methods implemented by multiple expert reviewers to assess study quality. Furthermore, some difficulties in interpreting or a low rating occurred when information and methodological details were lacking in the studies. This highlights the necessity to increase the quality and the standard of reporting. As such, the COSMIN can be used both as a post hoc assessment tool and as a guideline to ensure that the essential information is reported for a proper evaluation of the psychometric properties and methodological quality of studies.

Based on the literature reviews and the selected instruments,¹³ the constructs/domains assessed by the PROMs were symptoms, pain, sport and recreational function, participation in physical activity, activity of daily living, physical function, and quality of life. A previous study reported that pain and fear of the condition worsening are the two main reasons to undergo surgery in patients with FAI⁵⁹ together with improvement of everyday life and the ability to do sport. Most PROMs proposed for patients with hip-related pain include these domains. However, other constructs and transition questions such as satisfaction and patient acceptable symptoms state that were not addressed in this consensus may be important.

Conclusions

Although not all the psychometric properties can be considered adequate, the participants of the first International Hip-related Pain Research Network consensus meeting recommend the HAGOS and iHOT for use in young and middle aged active adults with hip-related pain. The participants agreed that generic quality of life measures such as EQ-5D and SF-36 may be a useful addition. Nevertheless, more methodological studies are needed to further evaluate the validity of these instruments and the others excluded from the recommended PROMs.

Conflict of interest

Two members of the panel (DG, KT) belong to the research groups developing the instruments recommended in the current consensus. However, they did not take part in the literature update and the quality scoring of the studies and the psychometric properties.

Figure legends

Figure 1. PRISMA flow chart for the systematic review of the systematic reviews

Figure 2. PRISMA flow chart for the literature update

Figure 3: Consensus voting on statements. S=statement

Table captions

Table 1. Rating of the quality of the systematic reviews according to the AMSTAR2

Table 2. Rescoring of the articles from the review of Thorborg et al. (2015)

Table 3. Scoring of the articles from the review update

Table 4. Quality of the psychometric properties (overall rating and quality of the evidence) according to the COSMIN guidelines

Table 5. Summary results for content validity

Table 6: The final recommendations voted on at the consensus meeting, and results of the consensus voting

Appendix captions

Appendix 1. (Appendix 1_Summary SR and AMSTAR2.xlsx)

Appendix 3. Appendix 3_Critical COSMIN issues rating study quality.docx

Appendix 2. (Appendix 2_SummaryRatingPROMcharacteristics.xlsx)

Appendix 4. Appendix 4_DevelopmentPapers_ContentRating.xlsx

References

1. Matthews RD, Neumayer L. Inguinal hernia in the 21st century: an evidence-based review. *Curr Probl Surg* 2008;45(4):261-312. doi: 10.1067/j.cpsurg.2008.01.002 [published Online First: 2008/03/25]
2. Thorborg K, Rathleff MS, Petersen P, et al. Prevalence and severity of hip and groin pain in sub-elite male football: a cross-sectional cohort study of 695 players. *Scand J Med Sci Sports* 2017;27(1):107-14. doi: 10.1111/sms.12623 [published Online First: 2015/12/09]
3. van der Waal JM, Bot SD, Terwee CB, et al. The course and prognosis of hip complaints in general practice. *Ann Behav Med* 2006;31(3):297-308. doi: 10.1207/s15324796abm3103_12 [published Online First: 2006/05/17]
4. Jansen JA, Mens JM, Backx FJ, et al. Treatment of longstanding groin pain in athletes: a systematic review. *Scand J Med Sci Sports* 2008;18(3):263-74. doi: 10.1111/j.1600-0838.2008.00790.x [published Online First: 2008/04/10]
5. Vos T, Flaxman AD, Naghavi M, et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2012;380(9859):2163-96. doi: 10.1016/S0140-6736(12)61729-2 [published Online First: 2012/12/19]
6. Kostanjsek N. Use of The International Classification of Functioning, Disability and Health (ICF) as a conceptual framework and common language for disability statistics and health information systems. *BMC Public Health* 2011;11 Suppl 4:S3. doi: 10.1186/1471-2458-11-S4-S3 [published Online First: 2011/06/03]
7. Reiman MP, Peters S, Sylvain J, et al. Prevalence and Consistency in Surgical Outcome Reporting for Femoroacetabular Impingement Syndrome: A Scoping Review. *Arthroscopy* 2018;34(4):1319-28 e9. doi: 10.1016/j.arthro.2017.11.037 [published Online First: 2018/02/07]
8. Patrick DL, Burke LB, Powers JH, et al. Patient-reported outcomes to support medical product labeling claims: FDA perspective. *Value Health* 2007;10 Suppl 2:S125-37. doi: 10.1111/j.1524-4733.2007.00275.x [published Online First: 2007/11/13]
9. Group M. Patient-Reported Outcomes in Orthopaedics. *J Bone Joint Surg Am* 2018;100(5):436-42. doi: 10.2106/JBJS.17.00608 [published Online First: 2018/03/07]
10. Pugely AJ, Martin CT, Harwood J, et al. Database and Registry Research in Orthopaedic Surgery: Part 2: Clinical Registry Data. *J Bone Joint Surg Am* 2015;97(21):1799-808. doi: 10.2106/JBJS.O.00134 [published Online First: 2015/11/06]
11. Gagnier JJ. Patient reported outcomes in orthopaedics. *J Orthop Res* 2017;35(10):2098-108. doi: 10.1002/jor.23604 [published Online First: 2017/05/18]
12. Marshall M, Lockwood A, Bradley C, et al. Unpublished rating scales: a major source of bias in randomised controlled trials of treatments for schizophrenia. *Br J Psychiatry* 2000;176:249-52. [published Online First: 2001/02/07]
13. Thorborg K, Tjissen M, Habets B, et al. Patient-Reported Outcome (PRO) questionnaires for young to middle-aged adults with hip and groin disability: a systematic review of the clinimetric evidence. *Br J Sports Med* 2015;49(12):812. doi: 10.1136/bjsports-2014-094224 [published Online First: 2015/01/15]
14. Bhangu A, Singh P, Pinkney T, et al. A detailed analysis of outcome reporting from randomised controlled trials and meta-analyses of inguinal hernia repair. *Hernia* 2015;19(1):65-75. doi: 10.1007/s10029-014-1299-4 [published Online First: 2014/08/13]
15. Banerjee A, Hendrick P, Bhattacharjee P, et al. A systematic review of outcome measures utilised to assess self-management in clinical trials in patients with chronic pain. *Patient Educ Couns* 2018;101(5):767-78. doi: 10.1016/j.pec.2017.12.002 [published Online First: 2017/12/21]
16. Alviar MJ, Olver J, Brand C, et al. Do patient-reported outcome measures in hip and knee arthroplasty rehabilitation have robust measurement attributes? A systematic review.

Journal of rehabilitation medicine 2011;43(7):572-83. doi: 10.2340/16501977-0828
[published Online First: 2011/05/25]

17. d'Entremont AG, Cooper AP, Johari A, et al. What clinimetric evidence exists for using hip-specific patient-reported outcome measures in pediatric hip impingement? *Clin Orthop Relat Res* 2015;473(4):1361-7. doi: 10.1007/s11999-014-4027-6 [published Online First: 2014/11/05]
18. Gagnier JJ, Huang H, Mullins M, et al. Measurement Properties of Patient-Reported Outcome Measures Used in Patients Undergoing Total Hip Arthroplasty: A Systematic Review. *JBJS reviews* 2018;6(1):e2. doi: 10.2106/jbjs.Rvw.17.00038 [published Online First: 2018/01/10]
19. Harris K, Dawson J, Gibbons E, et al. Systematic review of measurement properties of patient-reported outcome measures used in patients undergoing hip and knee arthroplasty. *Patient Relat Outcome Meas* 2016;7:101-8. doi: 10.2147/prom.S97774 [published Online First: 2016/08/16]
20. Lodhia P, Slobogean GP, Noonan VK, et al. Patient-reported outcome instruments for femoroacetabular impingement and hip labral pathology: a systematic review of the clinimetric evidence. *Arthroscopy* 2011;27(2):279-86. doi: 10.1016/j.arthro.2010.08.002 [published Online First: 2010/11/03]
21. Saleh KJ, Davis A. Measures for Pain and Function Assessments for Patients With Osteoarthritis. *J Am Acad Orthop Surg* 2016;24(11):e148-e62. doi: 10.5435/jaaos-d-16-00303
22. Strickland LH, Hamilton TW, Jenkinson CC, et al. Patient-Reported Outcome Measure for Early Postoperative Recovery Following Lower Limb Arthroplasty: A Systematic Review. *J Arthroplasty* 2016;31(12):2933-40. doi: 10.1016/j.arth.2016.06.026
23. Terwee CB, Bouwmeester W, van Elsland SL, et al. Instruments to assess physical activity in patients with osteoarthritis of the hip or knee: a systematic review of measurement properties. *Osteoarthritis Cartilage* 2011;19(6):620-33. doi: 10.1016/j.joca.2011.01.002 [published Online First: 2011/01/22]
24. Thorborg K, Roos EM, Bartels EM, et al. Validity, reliability and responsiveness of patient-reported outcome questionnaires when assessing hip and groin disability: a systematic review. *Br J Sports Med* 2010;44(16):1186-96. doi: 10.1136/bjsm.2009.060889 [published Online First: 2009/08/12]
25. Tijssen M, van Cingel R, van Melick N, et al. Patient-Reported Outcome questionnaires for hip arthroscopy: a systematic review of the psychometric evidence. *BMC Musculoskelet Disord* 2011;12:117. doi: 10.1186/1471-2474-12-117 [published Online First: 2011/05/31]
26. Veenhof C, Bijlsma JW, van den Ende CH, et al. Psychometric evaluation of osteoarthritis questionnaires: a systematic review of the literature. *Arthritis Rheum* 2006;55(3):480-92. doi: 10.1002/art.22001 [published Online First: 2006/06/02]
27. Baumann F, Popp D, Muller K, et al. Validation of a German version of the International Hip Outcome Tool 12 (iHOT12) according to the COSMIN checklist. *Health Qual Life Outcomes* 2016;14 doi: 10.1186/s12955-016-0407-9
28. Baumann F, Weber J, Zeman F, et al. Validation of a German version of the International Hip Outcome Tool (G-iHOT33) according to the COSMIN checklist: how much improvement is clinically relevant? *Arch Orthop Trauma Surg* 2016;136(1):83-91. doi: 10.1007/s00402-015-2336-1
29. Brans E, de Graaf JS, Munzbrock AVE, et al. Cross-Cultural Adaptation and Validation of the Dutch Version of the Hip and Groin Outcome Score (HAGOS-NL). *PLoS one* 2016;11(1) doi: 10.1371/journal.pone.0148119
30. Celik D, Can C, Aslan Y, et al. Translation, cross-cultural adaptation, and validation of the Turkish version of the Harris Hip Score. *Hip Int* 2014;24(5):473-79. doi: 10.5301/hipint.5000146
31. Chahal J, Van Thiel GS, Mather RC, et al. The Patient Acceptable Symptomatic State for the Modified Harris Hip Score and Hip Outcome Score Among Patients Undergoing Surgical Treatment for Femoroacetabular Impingement. *Am J Sports Med* 2015;43(8):1844-49. doi: 10.1177/0363546515587739

32. Costa RMP, Cardinot TM, Mathias LNCC, et al. Validation of the Brazilian version of the Hip Outcome Score (HOS) questionnaire. *Adv Rheumatol* 2018;58(1):4. doi: <http://dx.doi.org/10.1186/s42358-018-0007-y>
33. de Oliveira LP, Moura Cardinot T, Nunes Carreras Del Castillo L, et al. Translation and cultural adaptation of the Hip Outcome Score to the Portuguese language. *Rev Bras Ortop* 2014;49(3):297-304. doi: <https://dx.doi.org/10.1016/j.rboe.2014.03.011>
34. Giezen H, Stevens M, van den Akker-Scheek I, et al. Validity and reliability of the Dutch version of the Copenhagen Hip And Groin Outcome Score (HAGOS-NL) in patients with hip pathology. *PLoS one* 2017;12(10) doi: 10.1371/journal.pone.0186064
35. Hinman RS, Dobson F, Takla A, et al. Which is the most useful patient-reported outcome in femoroacetabular impingement? Test-retest reliability of six questionnaires. *Br J Sports Med* 2014;48(6):458-U153. doi: 10.1136/bjsports-2012-092072
36. Impellizzeri FM, Mannion AF, Naal FD, et al. Validation of the Core Outcome Measures Index in Patients With Femoroacetabular Impingement. *Arthroscopy* 2015;31(7):1238-46. doi: 10.1016/j.arthro.2015.01.014
37. Impellizzeri FM, Mannion AF, Naal FD, et al. Validity, Reproducibility, and Responsiveness of the Oxford Hip Score in Patients Undergoing Surgery for Femoroacetabular Impingement. *Arthroscopy* 2015;31(1):42-50. doi: 10.1016/j.arthro.2014.07.022
38. Jonasson P, Baranto A, Karlsson J, et al. A standardised outcome measure of pain, symptoms and physical function in patients with hip and groin disability due to femoro-acetabular impingement: cross-cultural adaptation and validation of the international Hip Outcome Tool (iHOT12) in Swedish. *Surg Sports Traumatol Arthrosc* 2014;22(4):826-34. doi: 10.1007/s00167-013-2710-x
39. Lee YK, Ha YC, Martin RL, et al. Transcultural adaptation of the Korean version of the Hip Outcome Score. *Surg Sports Traumatol Arthrosc* 2015;23(11):3426-31. doi: 10.1007/s00167-014-2946-0
40. Polat G, Celik D, Cil H, et al. Evidence for reliability, validity and responsiveness of Turkish version of Hip Outcome Score. *Acta Orthop Traumatol Turc* 2017;51(4):319-24. doi: 10.1016/j.aott.2017.05.001
41. Ruiz-Iban MA, Seijas R, Sallent A, et al. The international Hip Outcome Tool-33 (iHOT-33): multicenter validation and translation to Spanish. *Health Qual Life Outcomes* 2015;13 doi: 10.1186/s12955-015-0255-z
42. Seijas R, Sallent A, Ruiz-Iban MA, et al. Validation of the Spanish version of the hip outcome score: a multicenter study. *Health Qual Life Outcomes* 2014;12 doi: 10.1186/1477-7525-12-70
43. Stevens M, van den Akker-Scheek I, ten Have B, et al. Validity and Reliability of the Dutch Version of the International Hip Outcome Tool (iHOT-12NL) in Patients With Disorders of the Hip. *J Orthop Sports Phys Ther* 2015;45(12):1026-2. doi: <http://dx.doi.org/10.2519/jospt.2015.6178>
44. Thomee R, Jonasson P, Thorborg K, et al. Cross-cultural adaptation to Swedish and validation of the Copenhagen Hip and Groin Outcome Score (HAGOS) for pain, symptoms and physical function in patients with hip and groin disability due to femoro-acetabular impingement. *Surg Sports Traumatol Arthrosc* 2014;22(4):835-42. doi: 10.1007/s00167-013-2721-7
45. Tijssen M, Tak I, Stubbe J, et al. Translation, Cross-cultural Adaptation, and Validation of the Dutch International Hip Outcome Tool-33 (iHOT-33 NL) in Young, Physically Active Individuals With Symptomatic Hip Joint Pathology. *J Orthop Sports Phys Ther* 2018;48(4):289-+. doi: 10.2519/jospt.2018.7610
46. Li DH, Wang W, Li X, et al. Development of a valid Simplified Chinese version of the International Hip Outcome Tool (SC-iHOT-33) in young patients having total hip arthroplasty. *Osteoarthritis Cartilage* 2017;25(1):94-98. doi: 10.1016/j.joca.2016.08.013 [published Online First: 2016/09/14]

47. Collins NJ, Barton CJ, van Middelkoop M, et al. 2018 Consensus statement on exercise therapy and physical interventions (orthoses, taping and manual therapy) to treat patellofemoral pain: recommendations from the 5th International Patellofemoral Pain Research Retreat, Gold Coast, Australia, 2017. *Br J Sports Med* 2018 doi: 10.1136/bjsports-2018-099397
48. Crossley KM, van Middelkoop M, Callaghan MJ, et al. 2016 Patellofemoral pain consensus statement from the 4th International Patellofemoral Pain Research Retreat, Manchester. Part 2: recommended physical interventions (exercise, taping, bracing, foot orthoses and combined interventions). *Br J Sports Med* 2016;50(14):844-52. doi: 10.1136/bjsports-2016-096268
49. Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. *Ann Intern Med* 1993;118(8):622-9. [published Online First: 1993/04/15]
50. Brooks R. EuroQol: the current state of play. *Health Policy* 1996;37(1):53-72. [published Online First: 1996/06/06]
51. EuroQol G. EuroQol--a new facility for the measurement of health-related quality of life. *Health Policy* 1990;16(3):199-208. [published Online First: 1990/11/05]
52. McHorney CA, Ware JE, Jr., Raczek AE. The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Med Care* 1993;31(3):247-63. [published Online First: 1993/03/01]
53. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992;30(6):473-83. [published Online First: 1992/06/11]
54. Petrou S, Hockley C. An investigation into the empirical validity of the EQ-5D and SF-6D based on hypothetical preferences in a general population. *Health Econ* 2005;14(11):1169-89. doi: 10.1002/hec.1006 [published Online First: 2005/06/09]
55. Thorborg K, Holmich P, Christensen R, et al. The Copenhagen Hip and Groin Outcome Score (HAGOS): development and validation according to the COSMIN checklist. *Br J Sports Med* 2011;45(6):478-91. doi: 10.1136/bjsm.2010.080937 [published Online First: 2011/04/12]
56. Nilsdotter AK, Lohmander LS, Klassbo M, et al. Hip disability and osteoarthritis outcome score (HOOS)--validity and responsiveness in total hip replacement. *BMC Musculoskelet Disord* 2003;4:10. doi: 10.1186/1471-2474-4-10 [published Online First: 2003/06/05]
57. Roos EM, Roos HP, Lohmander LS, et al. Knee Injury and Osteoarthritis Outcome Score (KOOS)--development of a self-administered outcome measure. *J Orthop Sports Phys Ther* 1998;28(2):88-96. doi: 10.2519/jospt.1998.28.2.88 [published Online First: 1998/08/12]
58. Collins NJ, Prinsen CA, Christensen R, et al. Knee Injury and Osteoarthritis Outcome Score (KOOS): systematic review and meta-analysis of measurement properties. *Osteoarthritis Cartilage* 2016;24(8):1317-29. doi: 10.1016/j.joca.2016.03.010 [published Online First: 2016/03/26]
59. Mannion AF, Impellizzeri FM, Naal FD, et al. Fulfilment of patient-rated expectations predicts the outcome of surgery for femoroacetabular impingement. *Osteoarthritis Cartilage* 2013;21(1):44-50. doi: 10.1016/j.joca.2012.09.013 [published Online First: 2012/10/17]

Table 1. Rating of the quality of the systematic reviews according to the AMSTAR2

First author	Research question and inclusion criteria	Review methods established prior to the conduct	Selection of study design explained*	Literature strategy	Study selection in duplicate	Data extraction in duplicate	List of excluded studies	Studies described in adequate detail	Risk of bias of individual studies	Source of funding for the studies included	Risk of bias of studies taken into account	Conflict of interest reported	Overall rating**
	1	2	3	4	5	6	7	8	9	10	13 [#]	16	
Veenof et al. 2006	Y	N	Y	N	Y	Y	N	Y	N	N	N	Y	CRITICALLY LOW
Thorborg et al. 2010	Y	N	Y	Y	Y	Y	N	Y	N	N	N	Y	CRITICALLY LOW
Lodhia et al. 2011	Y	N	Y	N	Y	Y	Y	N	N	N	N	Y	CRITICALLY LOW
Tijessen et al. 2011	Y	N	Y	Y	Y	Y	N	Y	Y	N	Y	Y	MODERATE
Thorborg et al. 2015	Y	Y	Y	Y	Y	Y	N	Y	Y	N	Y	Y	MODERATE
d'Entremont et al. 2015	Y	N	N	PY	Y	Y	N	Y	N	N	N	Y	CRITICALLY LOW

Note: items 11, 12, 14 and 15 were not applicable (see explanation in the text); *being systematic reviews of studies on psychometric properties, the design explanation has been interpreted as referring to methodological papers specifically addressing psychometric attributes; #, the risk of bias assessment for methodological papers was considered the use of established criteria such as the COnsensus-based Standards for the selection of health Measurement Instruments (COSMIN) checklist; Y=yes; N=no

Table 2. Rescoring of the articles from the review of Thorborg et al. (2015)

Studies (Authors, year)	Questionnaires	Structural validity	Internal consistency	Cross cultural validity	Reliability	Meas error	Construct validity	Responsiveness
<i>Papers selected from the review of Thorborg et al. 2015</i>								
Griffin et al. 2012	iHOT-12	Inadequate			Doubtful		Doubtful	Very good
Jonasson et al.	iHOT-12	Adequate	Inadequate	Adequate	Adequate	Adequate	Very good	Very good
Mothadi et al. 2012	iHOT-33	Inadequate	Inadequate		Adequate		Adequate	Adequate
Polesello et al. 2012	iHOT-33			Inadequate				
Kemp et al. 2013	iHOT-33-HOS-HAGOS		Doubtful		Adequate	Adequate	Very good	Very good
Hinman et al. 2014	iHOT-33-HOS-HAGOS			Inadequate	Adequate	Adequate		
Martin et al. 2006	HOS	Inadequate	Very good				Adequate	
Martin et al. 2007	HOS						Adequate	
Martin et al. 2008	HOS				Adequate	Adequate		Adequate
Naal et al. 2011	HOS		Very good	Inadequate	Adequate	Adequate	Very good	
Lee et al. 2014	HOS		Very good	Adequate	Doubtful		Very good	Inadequate
Polat et al. 2017	HOS		Very good	Adequate	Adequate	Adequate	Adequate	Adequate
Seijas et al. 2014	HOS		Very good	Adequate	Adequate	Adequate	Very good	Very good
Thorborg et al. 2011	HAGOS	Inadequate	Very good		Very good	Very good	Very good	Very good
Kemp et al. 2013	HAGOS		Very good	Adequate	Adequate	Adequate	Very good	Very good
Thomee et al. 2014	HAGOS	Inadequate	Very good	Very good	Adequate	Adequate	Very good	Very good

Table 3. Scoring of the articles from the review update

Studies (Authors, year)	Questionnaires	Structural validity	Internal consistency	Cross cultural validity	Reliability	Meas error	Construct validity	Responsiveness
<i>Papers from the update</i>								
Lee et al, 2015	HOS		Very good	Adequate	Doubtful		Very good	Inadequate
Ruiz-Iba et al 2015	iHOT-33		Doubtful	Very good	Adequate	Adequate	Inadequate	Adequate
Impellizzeri et al 2015	COMI						Adequate	Doubtful
Impellizzeri et al 2015	OHS	Adequate	Inadequate		Adequate	Adequate	Inadequate	Adequate
Baumann 2016	iHOT-12		Very good	Adequate	Doubtful	Doubtful	Adequate	Inadequate
Baumann 2016	iHOT-33		Doubtful	Adequate	Doubtful	Doubtful	Adequate	Inadequate
Polat et al. 2017	HOS		Very good	Adequate	Adequate	Adequate	Adequate	Adequate
Costa 2018	HOS		Very good	Doubtful	Doubtful	Doubtful	Very good	
Tijssen et al 2018	iHOT-33	Inadequate	Inadequate	Adequate	Very good	Very good	Very good	
Li et al. 2017	iHOT-33		Doubtful	Adequate	Adequate	Adequate	Adequate	Adequate

Table 4. Quality of the psychometric properties (overall rating and quality of the evidence) according to the COSMIN guidelines

Questionnaires	Structural validity		Internal consistency		Cross cultural validity		Reliability		Measurement error		Hypotheses testing - Construct validity		Responsiveness	
	Overall rating	Evidence (Quality)	Overall rating	Evidence (Quality)	Overall rating	Evidence (Quality)	Overall rating	Evidence (Quality)	Overall rating	Evidence (Quality)	Overall rating	Evidence (Quality)	Overall rating	Evidence (Quality)
HAGOS	?	Very low	+	High	+	High	+	High	-	High	+	High	+	High
iHOT-33	?	Very low	+	Very low	+	High	+	High	-	Moderate	+	High	+	High
iHOT-12	?	Low	+	Moderate	+	High	+	Moderate	-	Low	+	High	?	High
HOS	?	Very low	+	High	+	High	+	High	?	High	+	High	+	High

+, sufficient; -, insufficient; ±, inconsistent; ?, indeterminate

Table 5. Summary results for content validity

	HAGOS		iHOT-33	
	Development study	Rating of reviewers	Development study	Rating of reviewers
<i>Relevance</i>				
Rating of results	±	+	±	+
Quality of evidence		Low		Low
<i>Comprehensiveness</i>				
Rating of results	?	+	?	+
Quality of evidence		Low		Low
<i>Comprehensibility</i>				
Rating of results	+	+	+	+
Quality of evidence		Low		Low
<i>Content validity</i>				
Rating of results	±	+	±	+
Quality of evidence		Low		Low

+, sufficient; -, insufficient; ±, inconsistent; ?, indeterminate

Table 6: The final recommendations voted on at the consensus meeting, and results of the consensus voting

Recommendation statements	Median	IQR	Mode	Consensus voting result
S1. The HAGOS and iHOT instruments (long and reduced versions) are the most appropriate PROMs to use in young and middle-aged active adults with hip-related pain.	9	8-9	9	Appropriate
S2. HAGOS and iHOT were developed mainly in surgical context. More research is needed into their utility in a non-surgical treatment context.	9	8-9	9	Appropriate
S3. EQ-5D and SF-36 are generic quality of life measures that can supplement the hip-related measures, HAGOS and iHOT.	9	8-9	9	Appropriate
S4. Future research should include further analysis of content and structural validity, and the relationship between individual measurement error and the minimal clinically important change for the recommended PROMs.	9	8-9	9	Appropriate

HAGOS = Copenhagen Hip and Groin Outcome score; IHOT = international hip outcome tool; EQ-5D = Euroqol 5 dimensions questionnaire; SF-36 = short form 36 questionnaire; PROM = patient reported outcome measure





