Measurement properties of the HOOS-PS in revision total hip arthroplasty: a validation study on validity, interpretability, and responsiveness in 136 revision hip arthroplasty patients

Sanne KORBEE [1], Robin VAN KEMPEN [1], Remco VAN WENSEN [1], Marieke VAN DER STEEN [1,2], and Wai-Yan LIU [1,2]

---

[1] Department of Orthopedic Surgery & Trauma, Catharina Hospital, Eindhoven; [2] Department of Orthopedic Surgery & Trauma, Máxima MC, Eindhoven, The Netherlands
Correspondence: sannekorbee@hotmail.com

Background and purpose — To determine whether the Hip disability and Osteoarthritis Outcome Score-Physical function Short-form (HOOS-PS) is able to appropriately evaluate physical function in revision hip arthroplasty patients, this study assesses psychometric properties of the Dutch HOOS-PS in this patient population.

Patients and methods — We assessed psychometric properties of the HOOS-PS following the Consensus-based Standards for the selection of health Measurement Instruments (COSMIN) criteria. Content validity, including comprehensibility, comprehensiveness, and relevance of the items, was assessed using cognitive debriefing interviews in hip revision patients (n = 8) and orthopedic surgeons specialized in revision surgery (n = 7). Construct validity, responsiveness, and interpretability (floor/ceiling effects) were assessed in revision hip arthroplasty patients (baseline n = 136, follow-up n = 67). We formulated hypotheses a priori to assess construct validity and responsiveness using the EuroQol 5-Dimensions Health Questionnaire, Numeric Rating scale for pain, and Oxford Hip Score as comparators. All questionnaires were measured at baseline and 1 year postoperatively.

Results — We found insufficient content validity of the HOOS-PS, as relevance and comprehensibility of the items scored < 85% on the COSMIN criteria for revision hip arthroplasty patients. Construct validity was sufficient as all hypotheses were confirmed (≥ 75% COSMIN criteria). Interpretability was sufficient (< 15% COSMIN criteria) and responsiveness was insufficient (< 75% COSMIN criteria).

Interpretation — The Dutch HOOS-PS is not able to sufficiently evaluate physical function in revision hip arthroplasty patients. Minor changes in the items are needed for the HOOS-PS to become sufficiently content valid, because the HOOS-PS lacks relevant items and comprehensiveness.

The number of revision total hip arthroplasties (revision THAs) is expected to rise over the years, due to an increase in life expectancy, prevalence of obesity, and extended indications for THAs in younger patients (**1-3**). With this increase in number of revision THAs, an increase in indications and revision strategies is also anticipated. Therefore, there is a need to appropriately evaluate and compare the outcome of different types of revision THAs.

In the assessment of medical treatment outcomes, patient-reported outcome measures (PROMs) can be used (**4**). The Hip disability and Osteoarthritis Outcome Score-Physical function Short-form (HOOS-PS) is a commonly used PROM, and measures physical functioning with fewer items than the full-length questionnaire (**5**), thereby reducing the burden of the responder and administrative load (**5,6**). The HOOS-PS has been validated to evaluate primary hip arthroplasty patients (**5,6**). However, the HOOS-PS has not yet been validated in revision arthroplasty patients.

To determine whether the Dutch HOOS-PS is able to appropriately evaluate physical function in revision arthroplasty patients, we aimed to assess the psychometric properties (content validity, construct validity, interpretability, and responsiveness) of the Dutch HOOS-PS in a revision hip arthroplasty population.

Patients and methods

Participants

221 patients aged 18 years or older and who were receiving a revision THA were consecutively recruited from the department of orthopedics of Catharina Hospital (Eindhoven, The Netherlands) between March 2015 and June 2019. Patients

were not included if they had insufficient understanding of the Dutch language. An additional study population, consisting of 7 patients and 8 orthopedic surgeons, was recruited between January and April 2020 to participate in cognitive debriefing interviews. Directly after the indication was made for revision arthroplasty, patients were asked to participate in cognitive debriefing interviews at the outpatient department. The same inclusion and exclusion criteria were applied. Orthopedic surgeons were considered eligible when specialized in revision arthroplasty surgery. Recruitment of orthopedic surgeons took place through purposive sampling in 3 independent high-volume revision arthroplasty centers in the Netherlands: Catharina Hospital (Eindhoven), Máxima MC (Eindhoven/Veldhoven), and Sint Maartenskliniek (Nijmegen).

Data collection

Preoperatively (T0), patients were asked to complete the Dutch versions of the HOOS-PS, Oxford Hip Score (OHS), Numeric Rating Scale for pain (NRS-pain), and EuroQol 5-Dimensions Health Questionnaire (EQ-5D-3L). Data retrieved at T0 was used to assess the construct validity and interpretability of the HOOS-PS. 1 year after revision surgery (T1), patients were asked to complete the PROMs again, as well as the anchor question, "To what extent has your general daily functioning changed since the surgery?". The anchor question consisted of 7 response options, ranging from "very much deteriorated" to "very much improved." Data retrieved at T1 was used to assess the responsiveness of the HOOS-PS.

Furthermore, one-on-one cognitive debriefing interviews were conducted preoperatively to assess the content validity of the HOOS-PS. These interviews with 7 patients were consistently conducted, recorded, transcribed, and analyzed by 1 of the authors.

PROMS

HOOS-PS

This study investigates the psychometric properties of the Dutch HOOS-PS in a revision population. The HOOS-PS is a 5-item measure of physical function, intended to disclose difficulties in the patient's activities in the last week due to hip problems (Figure 1, see Supplementary data) (5). The questionnaire comprises selected items of daily living and function, sports, and recreational activity subscales of the original full version of the HOOS (7). The scoring system comprises a 5-point Likert scale on degree of difficulty, ranging from none to extreme difficulty with. The sum of the ordinal scores is converted to a Rasch-based 0–100 interval score, in which a higher score represents a lower degree of difficulty (5). The HOOS-PS was previously translated into Dutch via the forward–backward method (8).

Oxford Hip Score (OHS)

The OHS measures pain intensity and functional limitations during various activities of the hip joint (9). The OHS con-

sists of 12 items, which are subdivided into disease-specific and generic questions. Each item contains 5 answer options, ranging from 4 "no problems" to 0 "severe problems/unable to execute." The maximum total score is 48, which corresponds to the lowest pain intensity and the least functional hip limitation in the last 4 weeks. The questionnaire is validated for patients receiving primary and revision THA (10,11).

Numeric Rating Scale for pain (NRS-pain)

The NRS-pain scale assesses pain intensity during rest and movement (12). Patients are asked to score the level of pain they experience on a scale of 0 to 10, with 0 representing no pain and 10 representing the worst pain imaginable during movement. The NRS-pain is a valid questionnaire in adult patients with musculoskeletal-related problems (12).

EuroQol 5-Dimensions Health Questionnaire (EQ-5D-3L)

The EQ-5D-3L is a 5-dimensional questionnaire measuring the general health of the patient, including mobility, self-care, daily activities, pain/discomfort, and anxiety/depression (13). 3 answer options can be given to each dimension, which are: "no problems," "minor problems," or "severe problems/unable to execute." The index score ranges from 0 (representing death) to 1.0 (representing full health), with negative values representing states worse than death (13,14).

Statistics

Psychometric properties

The psychometric properties of the HOOS-PS were evaluated according to the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) criteria (15,16). Psychometric properties were evaluated in terms of content and construct validity, responsiveness, and interpretability.

Content validity

Content validity is the degree to which the content of the instrument is an adequate reflection of the construct to be measured (16,17). Content validity of the HOOS-PS was assessed by the cognitive debriefing interviews with 7 experts in revision arthroplasties and 8 patients receiving a revision THA, evaluating the comprehensibility, comprehensiveness, and relevance of the items. An item was defined as relevant if at least 85% of the interviewees scored it to be relevant for construct and target population. Subsequently, the HOOS-PS was defined as valid if a minimum of 85% of the items were scored to be relevant, 85% of the patients found the HOOS-PS comprehensive and comprehensible, and 85% of the experts found the HOOS-PS comprehensive (18).

Construct validity

Construct validity is the degree to which the scores of the instrument are consistent with hypotheses, regarding relationships to scores of comparative instruments and to differences

between groups, based on the assumption that the instrument validly measures the construct to be measured (4,16). There is no gold standard to assess the functional status of the joint after revision THA. Therefore, the construct validity was measured by the degree to which the scores of the HOOS-PS were equivalent to scores of other PROMs that aim to evaluate a similar construct and are already used or validated in assessing hip function. Construct validity was assessed using predefined hypotheses on the expected relationship between the HOOS-PS and comparative PROMs (OHS, NRS-pain, EQ-5D-3L). These hypotheses were formulated based on literature studies and expert opinions (19,20) and are shown in Table 3. Construct validity was evaluated by calculating Pearson's correlation coefficients between scores at baseline of the HOOS-PS, OHS, NRS-pain, and EQ-5D-3L, and patients' age, BMI, and Charnley classification. In the case of non-normally distributed data at baseline, a Spearman's correlation coefficient was calculated. All correlation coefficients were classified into 3 categories: high correlation ($r \geq 0.5$), moderate correlation ($r = 0.3–0.5$), and low correlation ($r \leq 0.3$), as recommended in the COSMIN criteria (4). These correlations were then compared with the predefined hypotheses. The construct validity of the HOOS-PS was considered sufficient if at least 75% of the predefined hypotheses met the results (4).

Responsiveness
Responsiveness is the ability of an instrument to detect change over time in the construct to be measured (16). We evaluated the responsiveness of the HOOS-PS by comparing the change in scores between T0 and T1 with the changes in its comparative PROMs. In addition, we compared the change in scores in patients whose answer to the anchor question was that they had improved, stable, or deteriorated postoperative joint function compared with baseline. There were 7 answer options to the anchor questions, namely "very much improved/deteriorated," "much improved/deteriorated," "little improved/deteriorated," and "no difference." The stable joint function group consisted of patients whose answer to the anchor question was that there was no difference or little improvement/deterioration. The improved and deteriorated joint function groups represented patients who answered much or very much improved or deteriorated, respectively.

We calculated the effect size (ES) as follows: ([mean T1–mean T0]/SD of T0). Furthermore, the standardized response mean (SRM) was calculated as follows: ([mean T1–mean T0]/SD of change). Hypotheses on the expected correlations between the PROMs, and on the ES and SRM were formulated a priori based on literature studies and expert opinion (19,20) (Table 5). Pearson's correlation coefficients were calculated and were classified into 3 categories: high correlation ($r \geq 0.5$), moderate correlation ($r = 0.3–0.5$), and low correlation ($r \leq 0.3$), as recommended in the COSMIN criteria (4). We considered an ES ≥ 0.2 a small effect, ES ≥ 0.5 a medium effect, and ES ≥ 0.8 a large effect (4). Responsiveness was

considered sufficient if a minimum of 75% of the predefined hypotheses were in line with the results (4).

Interpretability
Interpretability is the degree to which qualitative meaning can be assigned to a score (16). We assessed the interpretability by examining the distribution of the HOOS-PS T0 scores, including preoperative and postoperative mean, SD, and floor and ceiling effects. Floor and ceiling effects were considered present if $\geq 15\%$ of the patients scored 0–5 or 95–100, respectively. The interpretability of the HOOS-PS was considered sufficient if floor and ceiling effects were absent (21).

Analysis
Interviews for content validity were transcribed and labelled using ATLAS.ti (ATLAS.ti Scientific Software Development GmbH, Berlin, Germany). Data was analyzed using IBM SPSS statistics 25 (IBM Corp, Armonk, NY, USA). No data imputation was applied, and cases with missing data were excluded on analysis-by-analysis basis.

Ethics, data sharing, funding, and potential conflicts of interest

Results

221 patients fulfilled the inclusion criteria. However, 85 (38%) patients were excluded on analysis-by-analysis basis due to incomplete data on the questionnaires at T0. 136 (62%) patients completed all T0 PROMs, representing the baseline cohort. Due to missing responses and incomplete data on the questionnaires at T1, 69 (51%) patients were lost to follow-up. Consequently, 67 (49% of the baseline cohort) patients completed the T1 PROMs and answered the anchor question, representing the follow-up cohort (Figure 2). The baseline and follow-up cohort show no major systematic difference in baseline characteristics (Table 1).

Content validity (Table 2)
The interviewees defined 2 out of 5 items to be relevant for construct and target population. Patients reported having difficulty completing the HOOS-PS due to lack of relevance of the items "running" and "twisting/pivoting on loaded leg." The HOOS-PS was comprehensible to all patients. However, the interviewees reported having difficulty scoring the item "getting in/out of bath/shower" as 1 action. None of the interview-
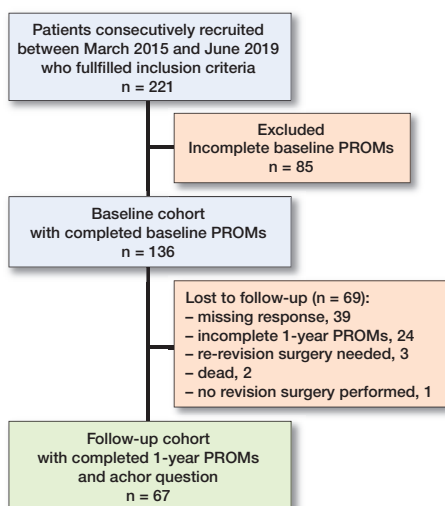
Figure 2. Number of patients included in analysis and reasons for loss to follow-up. PROMs: patient-reported outcome measures.

Table 2. Content validity: relevance, comprehensiveness, and comprehensibility of the HOOS-PS

Items	Experts' relevance (n = 7)	Patients' relevance (n = 8)	Total (n = 15)	Relevant (yes > 85%)
1. Descending stairs	6/7	8/8	14/15	Yes
2. Getting in/out of bath or shower	5/7	7/8	12/15	No
3. Sitting	7/7	8/8	15/15	Yes
4. Running	0/7	2/8	2/15	No
5. Twisting/pivoting on a loaded leg	4/7	6/8	10/15	No
Relevant items [a] (%)			2/5 (40)	
Comprehensiveness [b] (%)	0/7	0/8	0/15 (0)	
Comprehensibility [c] (%)		8/8	8/8 (100)	

[a] Proportion of experts and patients finding the items of the HOOS-PS relevant.
[b] Proportion of experts and patients finding the HOOS-PS comprehensive.
[c] Proportion of patients finding the HOOS-PS comprehensible.

Table 1. Baseline characteristics of revision total hip arthroplasty patients. Values are mean (SD) unless otherwise specified

Factor	Baseline cohort n = 136	Follow-up cohort n = 67
Baseline characteristics		
Female, n (%)	70 (52)	32 (48)
Age	74 (11) [a]	72(8.6)
BMI	26 (3.5)	26 (4.0) [a]
Baseline outcome measures		
HOOS-PS	55 (21)	55 (20)
OHS	26(10)	26 (9.7)
NRS-pain	6.4 (2.8)	6.6 (2.6)
EQ-5D-3L	0.51 (0.30)	0.57 (0.29)

[a] Median (IQR) is used.
HOOS-PS: Hip disability and Osteoarthritis Outcome Score-Physical function Short-form; OHS: Oxford Hip Score; NRS-pain: Numeric Rating Scale for pain; EQ-5D-3L: EuroQol 5-Dimensions Health Questionnaire.

Table 3. Construct validity: hypotheses and confirmation

Hypotheses	Pearson's correlation (CI)	Hypothesis confirmed
The correlation between HOOS-PS and		
OHS is ≥ 0.50 (high)	0.81 (0.71 to 0.91)	Yes
NRS-pain is ≤ –0.50 (high)	–0.71 (–0.83 to –0.59)	Yes
EQ-5D-3L is ≥ 0.50 (high)	0.58 (0.44 to 0.72)	Yes
Charnley score is ≤ 0.30 (low)	0.01 (–0.17 to 0.20)	Yes
patient's age is ≤ 0.30 (low)	–0.04 (–0.20 to 0.13) [a]	Yes
patient's BMI is ≤ 0.30 (low)	–0.02 (–0.19 to 0.16)	Yes
OHS is ≥ 0.10 higher than that between HOOS-PS and EQ-5D-3L		Yes
NRS-pain is ≥ 0.10 higher than that between HOOS-PS and EQ-5D-3L		Yes
Hypotheses confirmed (%)		8/8 (100)

[a] Spearman's correlation is used.
CI: 95% confidence interval
For abbreviations, see Table 1.

ees defined the HOOS-PS as comprehensive. Frequently mentioned missing items, by both patients and orthopedic revision surgeons, were "walking," "bending to the floor," and "putting on socks/stockings" in the assessment of hip function in the revision THA population. Lastly, patients reported having difficulty interpreting the 5 response options of the HOOS-PS on the degree of difficulty, ranging from none to extreme. Patients reported struggling in the differentiation between "mild" and "moderate" difficulty.

Construct validity (Table 3)

A 100% confirmation of the hypotheses was found for the HOOS-PS, which is more than the 75% confirmation needed for sufficient construct validity of the questionnaire.

Responsiveness (Table 4, see Supplementary data)

Of the 67 follow-up hip patients, 40 patients reported in response to the anchor question that they experienced an improvement in hip function compared with baseline, 18 patients reported no change in hip function, and 9 patients reported deteriorated postoperative hip function. A total of 53% of the hypotheses were confirmed, which is less than the 75% that was needed for sufficient responsiveness of the HOOS-PS. Notably, patients who reported deteriorated postoperative hip function in fact showed improvement in the HOOS-PS scores at T1. However, the deteriorated group did show less improvement (from T0 to T1) in HOOS-PS scores compared with the stable function group, which, in turn, improved less than the improved function group.

Table 5. Floor and ceiling effects of HOOS-PS baseline scores. Values are count (%)

HOOS-PS baseline score (n = 136)	n (%)
0–5 points	1 (0.7)
95–100 points	6 (4.4)
Total	7 (5.1)

Interpretability (Table 5)

At baseline, 0.7% of the hip patients scored ≤ 5 points, and 4.4% scored ≥ 95 points on the HOOS-PS. These percentages are below the threshold of 15% for floor and ceiling effects.

Discussion

This is the first study to evaluate psychometric properties of the HOOS-PS in revision hip arthroplasty patients. Insufficient content validity and responsiveness, and sufficient construct validity and interpretability were found.

The content validity of the HOOS-PS was found to be insufficient. A 40% relevance for the construct and target audience was found for the items of the HOOS-PS, which is less than the 85% relevance that is needed for sufficient content validity of the questionnaire. Additionally, patients reported having difficulty completing the HOOS-PS, resulting in incomplete data at T0 and T1 and loss to follow-up. This implies that the items did not fully represent the revision THA population. Both experts and patients defined the item "running" as least relevant for the target audience. In the development of the HOOS-PS, Davis et al. interpreted running to be the most difficult of the selected activities (5). Where primary arthroplasty offers an opportunity to fully regain functional ability, it is conceivable that running is no longer relevant for revision THA patients, in particular when considering the relatively older patient population and poorer health condition of revision THA patients (22).

None of the interviewees defined the HOOS-PS as comprehensive. Frequently mentioned missing items were "walking," "bending to the floor," and "putting on socks/stockings." Notably, these items were considered for inclusion in the early stage of development of the original HOOS-PS, but were rejected based on misfit criteria in primary THA patients (5). Both experts and patients appeared to have difficulty interpreting the item "getting in/out of bath/shower," possibly because the item covers 2 different movements. The experts and patients explained that getting in or out of a bath was considered relevant in assessing hip function, since this action involves hip flexion, whereas getting in or out of a shower was not considered relevant. This resulted in an overall "not relevant" score for this item. These activities may be reconsidered in order to adequately evaluate physical function in the revision THA

population. Furthermore, patients appeared to have difficulty interpreting the 5 response options of the HOOS-PS on the degree of difficulty, ranging from none to extreme. Patients reported struggling in the differentiation between "mild" and "moderate" difficulty. Additionally, the interpretation of the response options varied widely between patients, resulting in potential response bias. These findings are in line with the results of a recent meta-analysis described by Braaksma et al. (23), suggesting the HOOS-PS inadequately reflects physical functioning in patients receiving THA.

We have assessed construct validity by means of hypothesis testing rather than criterion validity, as no gold standard is available. This study showed sufficient construct validity for the HOOS-PS, with confirmation of all hypotheses. As hypothesized, correlation coefficients between the PROMs, measuring joint function, were higher than influencing factors such as Charnley score, age, and BMI. Additionally, OHS and NRS-pain, which specifically measure function and pain, showed a higher correlation coefficient with the HOOS-PS than the EQ-5D-3L, which also measures anxiety and depression. Therefore, correlations for similar constructs were higher than for dissimilar constructs. Sufficient construct validity of the HOOS-PS was previously also described in primary THA patients (6).

As interpretability is recognized as an important aspect of measurement instrument by the COSMIN Delphi study, though not a measurement property, interpretability of the HOOS-PS was assessed in our study (17). We observed no floor or ceiling effects in the HOOS-PS baseline scores, indicating sufficient interpretability. These findings are in line with the results described by Ornetti et al. in primary hip arthroplasty patients (24).

Multiple methods of assessing responsiveness are described in the updated COSMIN criteria (4,15). A PROM should not only measure changes in the measured construct, but should also measure the right amount of change. Therefore, a combination of methods is recommended to adequately assess responsiveness of a PROM (4). In our study, the following 2 methods were used to assess the responsiveness of the HOOS-PS. The ability to detect change in physical function of the HOOS-PS was compared with other comparative PROMs' ability to detect change in physical function. This method is called "hypotheses regarding relationships to scores of comparative instruments" (15). Additionally, we assessed whether the HOOS-PS was able to distinguish between different patient responses to the anchor question. In the COSMIN criteria, this method is called "hypotheses regarding relationships to differences between groups" (15). Consequently, 53% of all hypotheses were confirmed, indicating insufficient responsiveness of the HOOS-PS. However, when looking only at the responsiveness evaluated by comparing the change in HOOS-PS scores with the comparative PROMs, sufficient responsiveness was observed. Hereby, 80% of the hypotheses was confirmed, indicating an ability to detect change in physical function after revision THA of the HOOS-PS similar to its comparative

PROMs. Similar correlation trends between change in HOOS-PS, OHS, and NRS-pain scores were described by Tolk et al. in primary THA patients (25).

The HOOS-PS tends to be less able to distinguish between different patient responses on the anchor question. Only 40% of the hypotheses formulated for the different response groups were confirmed. Although we have combined 2 methods to assess responsiveness and an adequate number of patients were evaluated according to the COSMIN criteria, it should be noted that the number of patients in each function group (deteriorated, stable, or improved function) were not equal. Moreover, the deteriorated group showed less improvement in HOOS-PS scores compared with the stable function group, which, in turn, improved less than the improved function group. These results suggest that the HOOS-PS is less able to detect the magnitude of change in postoperative joint function compared with baseline as experienced by the patients. Another explanation might be that patients have trouble remembering their preoperative physical function after 1 year. Consequently, patients are less able to compare preoperative with postoperative physical function. Additionally, expectations regarding postoperative functional status may have interfered with the interpretation and memory of their postoperative recovery. An opportunity for future research may lie in investigating the effect of patients' expectations on the experienced postoperative functional outcomes. However, in particular our results regarding the deteriorated group (Table 4 in Supplementary data, pre-set hypothesis numbers 9 and 10) should be interpreted with some caution.

The strength of this study is the inclusion of a large sample size of revision hip arthroplasty patients, making the results applicable in practice. The sample size was based on the COSMIN criteria, aiming for a "very good" score for the content validity (≥ 7 interviewees), construct validity (≥ 100 patients), and responsiveness and interpretability (≥ 50 patients) analysis (15). Additionally, the psychometric properties were assessed and interpreted using the renowned and updated COSMIN criteria, which makes this study highly reproducible for validation of questionnaires in other languages (4,15,17).

This study has several limitations. First, the OHS, NRS-pain, and EQ-5D-3L are validated questionnaires in various patient populations. However, there is limited evidence of validation of these PROMs in the revision arthroplasty population. This results in difficulty in evaluating and interpreting outcomes within the revision arthroplasty population. Second, the formulation of hypotheses is an arbitrary procedure, and the percentage of confirmed hypotheses depends on the number of hypotheses. In this study, a consensus on the formulation and number of hypotheses was reached among a panel of experts (researchers and orthopedic revision surgeons). Two of the interviewed experts for content validity also participated in this panel. However, hypotheses were for construct validity and responsiveness. Overall results and information on con-

tent validity were not available for the panel of experts. Nonetheless, pre-set opinion and possible bias cannot completely be eliminated. Third, all hypotheses were considered equal in the confirmation of the construct validity and responsiveness, while some hypotheses with stronger correlations may have had more weight than others. Lastly, since the HOOS-PS cannot contain missing values to calculate the interval score, cases with missing data were excluded on analysis-by-analysis basis (5). This resulted in exclusion of 85 revision THA patients, causing potential confounding in our data. Additionally, 96 of the included patients were lost to follow-up for the responsiveness analysis due to incomplete data and missing responses postoperatively. However, according to the COSMIN criteria, a "very good" sample size was still reached to assess the psychometric properties of a PROM (15). Additionally, patients in the baseline group did not show major systematic difference in baseline characteristics as compared with the follow-up group (Table 1).

In conclusion, considering the inconsistent results on the different psychometric properties, it is questionable whether the current version of the Dutch HOOS-PS should be used to evaluate physical function in a revision THA population. However, with minor adjustments the HOOS-PS has the potential to become a valid instrument to assess physical function in the revision THA population. We suggest deletion or adjustment of the item "running," as this study showed low relevance for this activity. Potential alternatives are "walking" and "bending to the floor." Further research into the psychometric properties of the adjusted HOOS-PS is needed to develop a valid PROM to assess physical function in the revision THA population.

1. **Maloney W J.** National Joint Replacement Registries: has the time come? J Bone Joint Surg 2001; 83: 1582-5. doi: 10.2106/00004623-200110000-00020.

2. **Kurtz S, Ong K, Lau E, Mowat F, Halpern M.** Projections of primary and revision hip and knee arthroplasty in the United States from 2005 to 2030. J Bone Joint Surg 2007; 89: 780-5. doi: 10.2106/JBJS.F.00222.

3. **Patel A, Pavlou G, Mújica-Mota R, Toms A.** The epidemiology of revision total knee and hip arthroplasty in England and Wales: a comparative analysis with projections for the United States. A study using the National Joint Registry dataset. Bone Joint J 2015; 97: 1076-81. doi: 10.1302/0301-620X.97B8.35170.

4. **Mokkink L B, Prinsen C, Patrick D L, Alonso J, Bouter L M, de Vet H, et al.** COSMIN methodology for systematic reviews of patient-reported outcome measures (PROMs). User manual 2018; 78: 1. doi: 10.1007/s11136-018-1798-3.

5. **Davis A, Perruccio A, Canizares M, Tennant A, Hawker G, Conaghan P, et al.** The development of a short measure of physical function for hip OA HOOS-Physical Function Shortform (HOOS-PS): an OARSI/OMERACT initiative. Osteoarthritis Cartilage 2008; 16: 551-9. doi: 10.1016/j.joca.2007.12.016.

6. **Davis A M, Perruccio A V, Canizares M, Hawker G A, Roos E M, Maillefert J-F, et al.** Comparative, validity and responsiveness of the HOOS-PS and KOOS-PS to the WOMAC physical function subscale in total joint replacement for osteoarthritis. Osteoarthritis Cartilage 2009; 17: 843-7. doi: 10.1016/j.joca.2009.01.005.

7. **Nilsdotter A K, Lohmander L S, Klässbo M, Roos E M.** Hip disability and osteoarthritis outcome score (HOOS): validity and responsiveness in total hip replacement. BMC Musculoskeletal Disord 2003; 4: 1-8. doi: 10.1186/1471-2474-4-10.

8. **Beaton D E, Bombardier C, Guillemin F, Ferraz M B.** Guidelines for the process of cross-cultural adaptation of self-report measures. Spine 2000; 25: 3186-91. doi: 10.1097/00007632-200012150-00014.

9. **Dawson J, Fitzpatrick R, Carr A, Murray D.** Questionnaire on the perceptions of patients about total hip replacement. J Bone Joint Surg Br 1996; 78: 185-90.

10. **Dawson J, Fitzpatrick R, Murray D, Carr A.** Comparison of measures to assess outcomes in total hip replacement surgery. Quality Health Care 1996; 5: 81-88. doi: 10.1136/qshc.5.2.81.

11. **Field R, Cronin M, Singh P.** The Oxford hip scores for primary and revision hip replacement. J Bone Joint Surg Br 2005; 87: 618-22. doi: 10.1302/0301-620X.87B5.15390.

12. **Hawker G A, Mian S, Kendzerska T, French M.** Measures of adult pain: visual analog scale for pain (vas pain), numeric rating scale for pain (nrs pain), McGill pain questionnaire (mpq), short–form McGill pain questionnaire (sf–mpq), chronic pain grade scale (cpgs), short form–36 bodily pain scale (sf–36 bps), and measure of intermittent and constant osteoarthritis pain (icoap). Arthritis Care Rese 2011; 63: S240-52. doi: 10.1002/acr.20543.

13. **EuroQol.** EuroQol: a new facility for the measurement of health-related quality of life. Health Policy 1990; 16: 199-208. doi: 10.1016/0168-8510(90)90421-9.

14. **Lamers L M, McDonnell J, Stalmeier P F, Krabbe P F, Busschbach J J.** The Dutch tariff: results and arguments for an effective design for national EQ–5D valuation studies. Health Econ 2006; 15: 1121-32. doi: 10.1002/hec.1124.

15. **Mokkink L B, Prinsen C A, Patrick D L, Alonso J, Bouter L M, de Vet H C, et al.** COSMIN Study Design checklist for patient-reported outcome measurement instruments; 2019. Available from: https://www.cosmin.nl/wp-content/uploads/COSMIN-study-designing-checklist_final.pdf.

16. **Terwee C B, Bot S D, de Boer M R, van der Windt D A, Knol D L, Dekker J, et al.** Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007; 60: 34-42. doi: 10.1016/j.jclinepi.2006.03.012.

17. **Terwee C B, Prinsen C A, Chiarotto A, Westerman M, Patrick D L, Alonso J, et al.** COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. Qual Life Res 2018; 27: 1159-70. doi: 10.1007/s11136-018-1829-0.

18. **Terwee C B, Prinsen C, Chiarotto A, de Vet H, Bouter L M, Alonso J, et al.** COSMIN methodology for assessing the content validity of PROMs: user manual. Amsterdam, Netherlands: VU University Medical Center; 2018.

19. **Ruyssen-Witrand A, Fernandez-Lopez C, Gossec L, Anract P, Courpied J, Dougados M.** Psychometric properties of the OARSI/OMERACT osteoarthritis pain and functional impairment scales: ICOAP, KOOS-PS and HOOS-PS. Clin Exp Rheumatol 2011; 29: 231.

20. **Greimel F, Dittrich G, Schwarz T, Kaiser M, Krieg B, Zeman F, et al.** Course of pain after total hip arthroplasty within a standardized pain management concept: a prospective study examining influence, correlation, and outcome of postoperative pain on 103 consecutive patients. Arch Orthop Trauma Surg 2018; 138: 1639-45. doi: 10.1007/s00402-018-3014-x.

21. **Van Der Velden C A, Van Der Steen M, Leenders J, Van Douveren F Q, Janssen R P, Reijman M.** Pedi-IKDC or KOOS-child: which questionnaire should be used in children with knee disorders? BMC Musculoskelet Disord 2019; 20: 1-8. doi: 10.1186/s12891-019-2600-6.

22. **Healy W L, Iorio R, Lemos M J.** Athletic activity after joint replacement. Am J Sports Med 2001; 29: 377-88. doi: 10.1177/03635465010290032301.

23. **Braaksma C, Wolterbeek N, Veen M, Prinsen C, Ostelo R.** Systematic review and meta-analysis of measurement properties of the Hip disability and Osteoarthritis Outcome Score-Physical Function Shortform (HOOS-PS) and the Knee Injury and Osteoarthritis Outcome Score-Physical Function Shortform (KOOS-PS). Osteoarthritis Cartilage 2020. doi: 10.1016/j.joca.2020.08.004.

24. **Ornetti P, Perruccio A, Roos E, Lohmander L, Davis A, Maillefert J.** Psychometric properties of the French translation of the reduced KOOS and HOOS (KOOS-PS and HOOS-PS). Osteoarthritis and cartilage 2009; 17: 1604-1608. doi: 10.1016/j.joca.2009.06.007.

25. **Tolk J J, Janssen R P, Prinsen C A, van der Steen M C, Bierma Zeinstra S M, Reijman M.** Measurement properties of the OARSI core set of performance-based measures for hip osteoarthritis: a prospective cohort study on reliability, construct validity and responsiveness in 90 hip osteoarthritis patients. Acta Orthop 2019; 90: 15-20. doi: 10.1007/s00167-017-4789-y.

Supplementary data

> **INSTRUCTIONS:** This survey asks for your view about your hip. This information will help us keep track of how well you are able to perform different activities. Answer every question by ticking the appropriate box, only <u>one</u> box for each question. If you are unsure about how to answer a question, please give the best answer you can so that you answer all the questions.
>
> The following questions concern your level of function in performing usual daily activities and higher level activities. For each of the following activities, please indicate the degree of difficulty you have experienced in the **last week** due to your hip problem.
>
> 1. Descending stairs
>   None   Mild   Moderate   Severe   Extreme
>   ❑   ❑   ❑   ❑   ❑
>
> 2. Getting in/out of bath or shower
>   None   Mild   Moderate   Severe   Extreme
>   ❑   ❑   ❑   ❑   ❑
>
> 3. Sitting
>   None   Mild   Moderate   Severe   Extreme
>   ❑   ❑   ❑   ❑   ❑
>
> 4. Running
>   None   Mild   Moderate   Severe   Extreme
>   ❑   ❑   ❑   ❑   ❑
>
> 5. Twisting/pivoting on your loaded leg
>   None   Mild   Moderate   Severe   Extreme
>   ❑   ❑   ❑   ❑   ❑

Figure 1. The Hip disability and Osteoarthritis Outcome Score-Physical function Short-form (5).

Table 4. Responsiveness: hypotheses and confirmation

Hypotheses	ES/SRM/Pearson's correlation (CI)	Hypothesis confirmed
1. The ES in patients who reported improved hip function at T1 is expected to be ≥ 0.80	1.40	Yes
2. The SRM in patients who reported improved hip function at T1 is expected to be ≥ 0.80	1.36	Yes
3. The ES in patients who reported stable hip function at T1 is expected to be ≤ 0.20	0.79	No
4. The SRM in patients who reported stable hip function at T1 is expected to be ≤ 0.20	1.04	No
5. The ES in patients who reported deteriorated hip function at T1 is expected to be ≥ 0.50	0.24	No
6. The SRM in patients who reported deteriorated hip function at T1 is expected to be ≥ 0.50	0.22	No
7. The ES in patients who reported improved hip function at T1 is expected to be ≥ 0.20 larger than the ES in patients who reported stable hip function at T1		Yes
8. The SRM in patients who reported improved hip function at T1 is expected to be ≥ 0.20 larger than the SRM in patients who reported stable hip function at T1		Yes
9. The ES in patients who reported deteriorated hip function at T1 is expected to be ≥ 0.20 larger than the ES in patients who reported stable hip function at T1		No
10. The SRM in patients who reported deteriorated hip function at T1 is expected to be ≥ 0.20 larger than the SRM in patients who reported stable hip function at T1		No
11. The correlation between changes in T0 and T1 scores of HOOS-PS and OHS is ≥ 0.50 (high)	0.87 (0.79–1.05)	Yes
12. The correlation between changes in T0 and T1 scores of HOOS-PS and NRS-pain is ≤ −0.50 (high)	−0.74 (−0.96 to −0.62)	Yes
13. The correlation between changes in T0 and T1 scores of HOOS-PS and EQ-5D-3L is ≥ 0.50 (high)	0.66 (0.45–0.80)	Yes
14. The correlation between changes in T0 and T1 scores between HOOS-PS and OHS is ≥ 0.10 higher than that between HOOS-PS and EQ-5D-3L		Yes
15. The correlation between changes in T0 and T1 scores between HOOS-PS and NRS-pain is ≥ 0.10 higher than that between HOOS-PS and EQ-5D-3L		No
Hypotheses confirmed (%)		8/15 (53)

ES: effect size; SRM: standardized response mean; CI: 95% confidence interval; T0: baseline; T1: 1 year postoperatively.
For abbreviations, also see Table 1.