Journal of Hip Preservation Surgery, 2021, **8**, 298–304
DOI: https://doi.org/10.1093/jhps/hnab073
Advance access publication 7 October 2021
Research article

Validation of the Dutch version of the Hip Outcome Score; validity, reliability, and responsiveness in patients with femoroacetabular impingement syndrome

Maarten A. Röling[1*], Brechtje Hesseling[2,3], Sebastiaan P.L. Jansen[4], Rolf M. Bloem[2,3] and Nina M.C. Mathijssen[2,3]

[1]Department of Orthopaedic Surgery, Reinier de Graaf Hospital Delft, Gelre Hospital, Gelre Apeldoorn Albert, Schweitzerlaan 31, Apeldoorn 7334 DZ, the Netherlands, [2]Reinier Haga Orthopaedic Center, Department of Orthopaedic Surgery, Toneellaan 2, Delft 2625 AD, the Netherlands, [3]Reinier de Graaf Hospital Delft, Reinier de Graafweg 5, Delft 2625 AD, the Netherlands and [4]Department of Orthopaedic Surgery, Alrijne Hospital, Simon Smitweg 1, Leiderdorp 2353 GA, the Netherlands

*Correspondence to: Maarten A. Röling E-mail: m.roling@gelre.nl

ABSTRACT

Due to a lack of a validated Dutch version of the Hip Outcome Score (HOS) considering functional outcome after hip arthroscopy for femoroacetabular impingement syndrome, we validated the Dutch version of the HOS (HOS-NL) in patients with femoroacetabular impingement syndrome for reliability, internal consistency, construct- and content validity. Furthermore, the smallest detectable change (SDC) and minimal clinically important difference (MCID) were determined. All consecutive patients scheduled for an arthroscopic procedure for FAIS were selected. Five questionnaires covering groin and hip pain were filled in at three moments in time (two pre-operatively with a maximum two-week interval and 6 months postoperatively). Main endpoints were reliability (test re-test, SDC), internal consistency (Cronbach alpha), construct validity (construct validity was considered sufficient if a least 75% of a-priori made hypotheses were confirmed), content validity (floor and ceiling effects) and responsiveness (MCID). The intraclass correlation coefficient (ICC) was 0.86 for the HOS ADL-NL and 0.81 for the HOS Sports-NL. SDC for the HOS ADL-NL was 21 and for the HOS Sports-NL 29 Cronbach alpha score was 0.882 for HOS ADL-NL and 0.792 for HOS Sports-NL. Construct validity was considered sufficient since 91% of the hypotheses were confirmed. No floor effects were determined. A small ceiling effect was determined for the HOS AD-NL postoperatively. The MCID for HOS ADL-NL and HOS Sports-NL were 14 and 11.0, respectively. The HOS-NL is a reliable and valid patient reported outcome measure for measuring physical function and outcome in active and young patients with femoroacetabular impingement syndrome.

INTRODUCTION

Patient-reported outcome measures (PROMs) are increasingly being used to evaluate clinical outcomes in orthopedics [1]. More orthopedic assessment tools are used, and many are predominantly developed for elderly patients [2, 3] who were supposed to suffer more from orthopedic-related functional limitations like osteoarthritis. However, young and active patients with hip or groin pain can suffer from femoroacetabular impingement syndrome (FAIS) [4–5]. Over the last decade, hip arthroscopy has become a popular and successful procedure for the treatment of FAIS in adults and adolescents, both male and female population [6–12]. To measure the outcome and results of arthroscopic surgery for FAIS, questionnaires should focus on the activities of these patients since most of these patients are physically more active compared to patients suffering from osteoarthritis [13–15]. The Hip Outcome Score (HOS) is an example of an English-language questionnaire focused on activities and sports and is considered a valid tool for measuring function in individuals who have undergone hip arthroscopy [16–18]. The HOS was intended to measure self-reported functional status, i.e. items that related to activity and participation were included. Tijssen [1] recommended the HOS for evaluating patients after hip arthroscopy for FAIS in a review in 2011 and many authors have used the HOS to describe postoperative results after hip arthroscopy for FAIS [15–19]. The HOS is especially designed for FAIS since it has a Sports domain covering a unique type of questions considering sports activities in patients. The HOS scored very high on observer agreement, internal consistency, test–retest reliability, construct validity, interpretability, and measurement error [16, 17]. In concordance with the international growth in the number of hip arthroscopies performed for FAIS, an increasing amount of hip arthroscopies is also performed in the Netherlands. To measure functional outcomes after arthroscopy for FAIS, several

Submitted 4 July 2021; Revised 20 August 2021; revised version accepted 16 September 2021

Dutch PROMs are available, like the international Hip Outcome Tool (iHOT) 12 NL and the Hip and Groin Outcome Score (HAGOS), but also a validated Dutch translation of the HOS is desirable. If several PROMs can be combined to measure functional outcome after hip arthroscopy for FAIS, this is more accurate and less influenced by the flaws of just that one PROM. As stated by Kluzek *et al.* [20], collecting multiple PROMS over time may help to overcome the single measure variability. The HOS is not yet translated into the Dutch language, nor is it validated for the Dutch language. We, therefore, translated the HOS questionnaire into the Dutch language (HOS-NL) in concordance with other translation studies into Spanish, Korean, Portuguese and German [21–24]. The quality of a PROM can be determined by several measurement properties, as stated by the COSMIN taxonomy [25, 26]. These properties are the reliability (internal consistency and test–retest reliability), validity (content validity and construct validity) and responsiveness. The objective of this study was, therefore, to evaluate these properties of the Dutch version of the HOS questionnaire in patients with FAIS. The smallest detectable change (SDC) and minimal clinically important difference (MCID) were determined. Our hypothesis was that the HOS-NL is a reliable and valid PROM for measuring physical function outcomes in ADL and sports-related activities in active and young patients with FAIS.

MATERIALS AND METHODS

The study was performed in the orthopedic surgery department of two large peripheral hospitals in (Blinded), (hospitals blinded) and contained two phases: translation and investigation of reliability and validity.

The local medical ethical committee approved the study (blinded).

All participating patients signed a written informed consent after being informed about the study. The preoperative assessment, operative treatment, and postoperative rehabilitation for FAIS were according to the local protocol and did not interfere with study participation.

Study population consisted of all consecutive patients with FAIS, derived from the orthopedic outpatient department from the two participating hospitals. Inclusion criteria were age between 18–65 years, a physical and radiological examination that confirms FAIS without severe osteoarthritis (≥Tönnis grade 3) [9], conservative treatment of FAIS of at least 6 months and physical activity. Exclusion criteria were patients with prior surgery to the hip for FAIS, a pathological fracture of the hip or other metastatic pathology and patients not speaking the Dutch language or refusing to participate.

We aimed to include at least 100 patients, based on recommendations of the COSMIN guidelines and other authors [25–29].

Translation procedure

A Dutch translation was made using a forward/backward translation protocol according to the guidelines of cross-cultural adaptations [30]. Since no major cultural differences in lifestyle exist between the Dutch and English/American population, we assumed that large cultural adaptation of the questionnaire was not required. For this first phase, the English version was translated into a Dutch version by two Dutch native speakers who speak the English language fluently, one with medical knowledge and one without. Both translations were combined by the translators and a team of experts (consisting of an orthopedic surgeon, a resident orthopedic surgery, and a researcher). Two persons translated the Dutch version back into an English version: both speaking English (native) as well as Dutch fluently. The final version was made by the research team. This version was tested in 20 patients with various hip pathologies (mainly FAIS) in the correct age category to determine whether the questions were understandable and whether patients were able to complete the questions. With these amendments, the final version was created as the HOS-NL.

Validation study

All participating patients completed several PROMs at three moments in time, twice before surgery with a maximum interval of 2 weeks and once at 6 months postoperatively. Patients completed the HOS-NL and translated versions of the modified Harris Hip Score (mHHS), the HAGOS-NL, the iHOT-12 NL and the numeric rating scale (NRS) for pain. Patients were asked to rate their own level of functioning due to their hip problems ('normal', 'almost normal', 'abnormal' or 'severely abnormal'), as well as the change in functioning after surgery ('much improved', 'somewhat improved', 'slightly improved', 'unchanged', 'slightly worse', 'somewhat worse' or 'much worse').

Reliability is defined as the ability of a test to yield the same results on repeated moments under the same conditions [31]. We used a 2-week interval preoperatively to define this. The test–retest reliability was assessed using the intraclass correlation coefficient (ICC) between the first and second applications of the HOS-NL. Values <0.5, between 0.5 and 0.75, between 0.75 and 0.90 and >0.90 were considered poor, moderate, good and excellent, respectively [32].

The measurement error is a combination of systematic and random error of scores in the HOS-NL, which is not determined by true change in the measured construct. To quantify the measurement error, we calculated the SDC. Data from T1 and T2 were used to determine the measurement error. We assumed that there would be no real change in patient's functioning within a 2-week interval, preoperatively.

Internal consistency is a measure based on the correlations between different items on the same test [26]. We used Cronbach's alpha [33]. A value exceeding 0.7 would indicate that the HOS-NL has good internal consistency in measuring functional outcome scores after surgery for FAIS [29].

Construct validity is the degree to which a test measures what it claims to be measuring [34]. The HOS-NL was therefore compared with the Dutch version of the HAGOS, the HAGOS-NL [35], the mHHS [36] and the Dutch version of the iHOT-12, the iHOT 12-NL [37–39], and the NRS for pain [40]. The association was determined by Pearson's correlation coefficients. Correlation coefficients can be considered small ($r < 0.30$), moderate ($r = 0.31–0.50$) or large ($r > 0.50$) or reversed ($r < -0.3$, $r = -0.3$ to -0.5, $r > -0.5$) when a maximum achievable score of one scale correlates with a minimum achievable score on the comparative scale [41]. If the instruments are measuring the same/similar attributes, the correlation coefficients should be between 0.4 and 0.8 [42]. A priori hypotheses were made concerning the correlations between the subscales. Construct validity was considered sufficient if at least 75% of the hypotheses were confirmed [43]. All hypotheses are summarized in Table IV.

Content validity addresses whether a questionnaire has enough items and adequately covers the domain of interest [53]. Content validity was evaluated by assessing the floor and ceiling effects of the questionnaire. Floor and ceiling effects were considered present if more than 15% of the respondents achieved the highest (95–100%) or lowest (0–5%) possible score [43].

Responsiveness is the ability of a measure to detect a change when an actual change has occurred, a change in response to a (surgical) intervention. To determine which change in HOS-NL scores can be interpreted as meaningful change, we calculated the MCID at 6 months postoperatively.

STATISTICS

Data were collected in Castor electronic database [44]. Statistical analyses were performed using IBM SPSS Version 22.0 for windows and Mac and in R using RStudio [45]. Patient characteristics were analyzed by means of descriptive statistics. *P*-values less than 0.05 were considered to indicate statistical significance.

The test–retest reliability was assessed using the ICC two-way mixed model [ICC(3,1), 95% CI] between the first and second applications of the HOS-NL. Paired *t*-tests were performed to determine the systematic difference between the first and second tests. R package 'psych' was used to calculate the ICC [46].

To calculate the SDC, we used the following formula: $\text{SDC} = 1.96 \times$ standard error of measurement (SEM) $\times \sqrt{2}$. SEM was calculated using the formula $\text{SEM} = \sqrt{\sigma^2_{\text{error}}}$, where σ^2_{error} is a variance component of the ICC [47].

To calculate the MCID, we used an anchor-based approach. The anchor question/criterion used to determine the MCID was whether patients reported being 'much improved', 'somewhat improved' or 'slightly improved' versus 'unchanged', 'slightly worse', 'somewhat worse' or 'much worse'. Based on sensitivity and specificity values, receiver operating characteristic curves were constructed for possible HOS change scores using R package 'pROC' [48]. Youden's cutoff was used to determine the MCID.

RESULTS

Patients were included from August 2017 to August 2020

Pretesting of the translated version of the HOS did not reveal any obstacles or any major difficulties for implementing and using the questionnaire. The HOS-NL version is added to the manuscript as a supplement.

A total of 135 patients were included for this study. A total of 111 patients had complete data (Fig. 1). Demographic characteristics are presented in Table I and the baseline and outcome scores of all PROMs are displayed in Table II.

The test–retest reliability of the HOS-NL subdomains based on calculated ICC values was good. The ICC values for the test–retest reliability are presented in Table III. SDC was 21 for

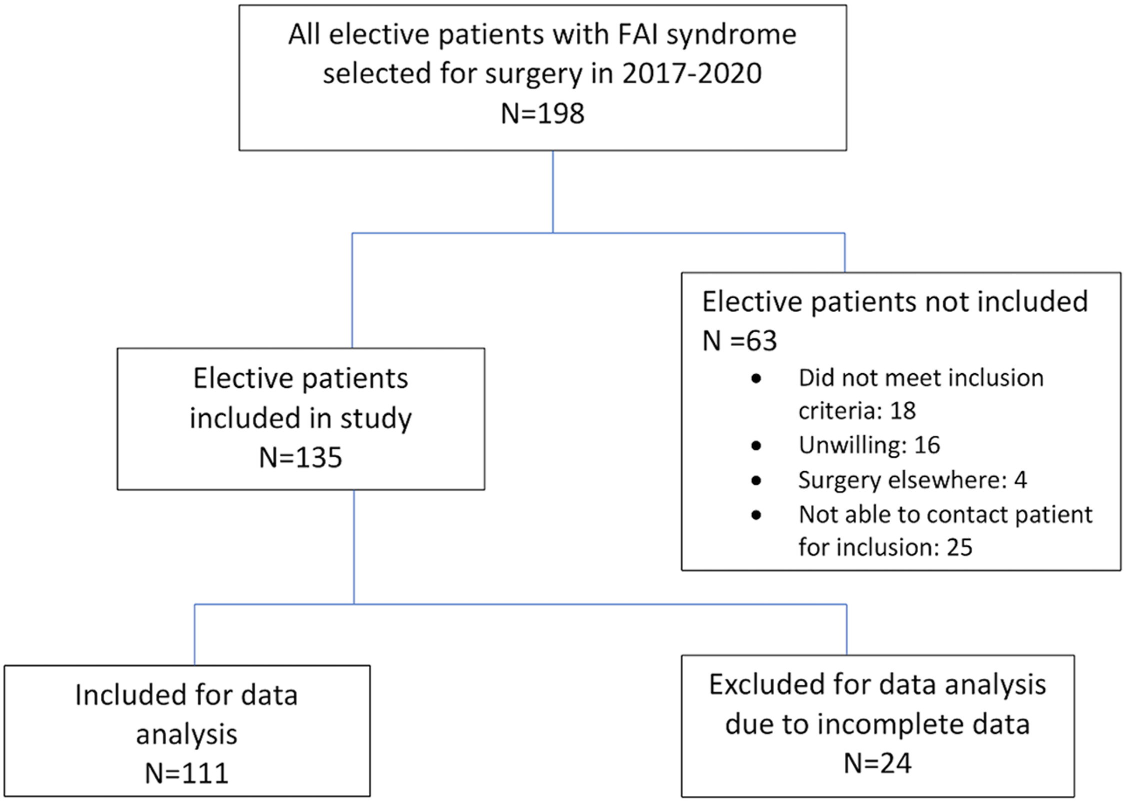
All elective patients with FAI syndrome selected for surgery in 2017-2020
N=198
Elective patients included in study
N=135
Elective patients not included
N =63
Did not meet inclusion criteria: 18
Unwilling: 16
Surgery elsewhere: 4
Not able to contact patient for inclusion: 25
Included for data analysis
N=111
Excluded for data analysis due to incomplete data
N=24

Fig. 1. Flow chart for inclusion.

the HOS ADL-NL and 29 for the HOS Sports-NL. Internal consistency was determined by Cronbach's alpha, which was 0.882 for the HOS ADL-NL and 0.792 for the HOS Sports-NL, which indicates a high level of internal consistency. The construct validity is considered sufficient because 91% of the hypotheses were confirmed. Table IV contains all correlations for this construct validity.

Content validity is presented in Table V, with the percentage of patients that scored 5% lowest possible and highest possible score: the floor and ceiling effects. No floor effect could be identified. A small ceiling effect was identified for the HOS-ADL NL postoperatively.

The responsiveness was determined by the MCID, presented in Table VI. For HOS ADL-NL, the MCID was 14 and for the HOS Sports-NL 11. The MCID is smaller than the SDC for both domains. Area under the curve is presented in Fig. 2.

Table I. Demographic characteristics

	N = *111*
Gender	M = 41 (37%) F = 70 (63%)
Mean age (range)	37.6 (18–59) SD 9.9
(American Society of Anesthesiologist physical status classification)	83 (75%)
ASA 1	
ASA 2	25 (22.5%)
ASA 3	3 (2.7%)
Affected side	Left 45 (40.5%) Right 66 (59.5%)
Diagnosis preoperative	
Cam	24 (22%)
Pincer	8 (7%)
Combined cam and pincer	5 (5%)
Labral tear	85 (77%)
Labral tear and FAI	15 (14%)
Other	3 (3%)

Table II. PROM scores of HOS ADL-NL, HOS Sports-NL, iHOT 12-NL, HAGOS ADL-NL, mHHS, and NRS for pain

	Preoperative score (SD)	*Postoperative score (SD) at 6 months*	*P*-value[b]
HOS ADL-NL	60.0 (19.0)	76.5 (20.8)	<0.001
HOS Sport-NL	41.2 (23.1)	61.6 (27.7)	<0.001
mHHS	39.1 (7.8)	43.7 (8.1)	<0.001
iHOT 12-NL	37.0 (17.6)	59.6 (25.6)	0.01
HAGOS ADL-NL	48.8 (24.9)	31.9 (27.3)	<0.001
NRS pain rest[a]	50.8 (25.4)	30.1 (29.1)	<0.001
NRS pain active[a]	68.4 (21.9)	44.1 (29.7)	<0.001

[a] NRS for pain on a visual analogue scale from 0 to 100.

[b] Differences between preoperative and postoperative PROM means were analyzed by independent Student *t*-test.

DISCUSSION

The results of this study offer evidence for test–retest reliability, validity, and responsiveness of the HOS-NL in young active individuals undergoing hip arthroscopic surgery for FAIS. This study also presents values to interpret change in scores over time, with SDC values of 21 and 29 over a 2-week preoperative period, and MCID values of 14 and 11 over a 6-month postoperative period for the HOS ADL-NL and Sport-NL, respectively.

The HOS is an important functional outcome tool that is used internationally to measure functional outcome after hip surgery [1]. Such a PROM must be validated for its purpose: i.e. testing functional outcome and changes in outcome [30]. It is, therefore, important to have validated these PROMs in patients' native language, in this case, the Dutch language.

Construct validity was determined by predefined hypotheses between the HOS-NL and other questionnaires. A minimum of 75% had to be confirmed to become a good construct validity [43]. Our hypotheses were confirmed in 91%. These correlations with other PROMs, such as the iHOT-12 and the NRS for pain, are comparable to other validation studies of the HOS [21–24]. The HOS-Brazil was validated in 2018 and showed high correlations with the Short-Form 12 and the Non-Arthritic Hip Score [23]. A Spanish version of the HOS was translated and validated in 2014 and showed equal correlation scores to the Western Ontario and McMaster Universities Osteoarthritis Index [21]. All validation studies showed good validity and internal consistency comparable with our results. Expected weaker correlations were found with HOS Sports-NL subscale and the HAGOS-NL and NRS. This weak correlation can be explained by the lack of specific sports-related scales in the HAGOS-NL or the NRS for pain. It is, therefore, difficult to compare the HOS Sports-NL to other questionnaires. This lack of specific sports PROMs is highlighted by a review of available PROMs in sports in 2019 [49], which concludes that there is a void in PROMs to evaluate the postoperative outcomes regarding the physical and psychological demands of athletes and sports practitioners. We think that the sports-related domain of the HOS is of additional value in this young and active patient population.

We determined a small ceiling effect in the HOS-ADL in 2% of all patients before surgery, which increased to 19% 6 months postoperatively. A ceiling effect in the HOS-Sports also developed during follow-up in 7.3%. Floor and ceiling effects might influence the reliability and validity if these effects occur in >15% of patients [18]. Thus, we can conclude that the ceiling effect in our analyses for the HOS ADL-NL postoperatively might influence the validity negatively.

The MCID is defined as the smallest measured change score that patients feel is important. If the MCID is smaller than the SDC, that clinically relevant change in score could not be safely detected above measurement error [50]. The MCID for the HOS is described by several authors. Nwachukwu [51] e.g. calculated an MCID of 8.8 at 1 year for the HOS ADL and 13.9 for the HOS Sports. Martin [16] has a different MCID for ADL and for Sports, 9 and 6, respectively. Ueland *et al.* [52] summarized these differences in a recent review in 2021. We determined an MCID of 14 for the HOS ADL-NL and 11 for HOS Sports-NL, which differed from the results of Martin [16] and Nwachukwu [51]. Differences in MCID between studies can be explained by

Table III. Test–retest reliability measures of HOS-ADL NL

	First measurement mean score (SD) T1	*Second measurement[a] mean score (SD) T2*	*ICC (R)[b]*	*Mean difference T1-T2 (95% CI)*
HOS ADL-NL	60.1 (19.6)	57.5 (21.0)	0.86	3.12 (0.94–5.29)
HOS Sports-NL	41.2 (24.0)	38.3 (24.5)	0.81	3.11 (0.12–6.10)

[a] <2 weeks after first measurement: mean time 11 days, standard deviation (SD) 6.3, 95% confidence interval (CI; 9.36–11.75).
[b] Intraclass correlation coefficient.

Table IV. Construct validity for HOS-NL

Subscale	*Questionnaire*	*Hypothesized correlation[b]*	*Calculated correlation T1[c]*	*Calculated correlation T3[d]*
HOS ADL-NL	HAGOS-NL ADL[a]	$r > 0.5$	$r = 0.826$	$r = 0.911$
HOS ADL-NL	HAGOS-NL QOL[a]	$r > 0.5$	$r = 0.589$	$r = 0.722$
HOS ADL-NL	HAGOS-NL S[a]	$r > 0.5$	$r = 0.670$	$r = 0.824$
HOS ADL-NL	iHOT 12-NL	$r > 0.5$	$r = 0.703$	$r = 0.839$
HOS ADL-NL	NRS pain	$r > -0.5$	$\boldsymbol{r = -0.486}$	$r = -0.550$
HOS Sports-NL	HAGOS-NL SR[a]	$r > 0.5$	$r = 0.797$	$r = 0.876$
HOS Sports-NL	NRS pain	$r > -0.5$	$\boldsymbol{r = -0.423}$	$r = -0.589$
HOS Sports-NL	HAGOS-NL QOL[a]	$r > 0.4$	$r = 0.597$	$r = 0.768$
HOS Sports-NL	HAGOS-NL S[a]	$r > 0.4$	$r = 0.600$	$r = 0.744$
HOS Sports-NL	iHOT 12-NL	$r > 0.3$	$r = 0.711$	$r = 0.819$
HOS Sports-NL	mHHS	$r > 0.3$	$r = 0.607$	$r = 0.718$

[a] Subdomains ADL, Quality of Life (QOL), Symptoms (S), and Sports and Recreation (SR).
[b] Determined with Pearson's correlation coefficient.
[c] T1: preoperatively.
[d] T3: 6 months postoperatively.
The incorrect hypothesized correlations are highlighted.

Table V. Content validity of HOS ADL-NL and HOS Sports-NL

	Preoperative (T1) floor effect N (%)	*Preoperative (T1) ceiling effect N (%)*	*6 Months postoperative (T3) floor effect (%)*	*6 Months postoperative (T3) ceiling effect (%)*
HOS ADL-NL	$N = 0$ (0%)	$N = 2$ (1%)	$N = 0$ (0%)	$N = 21$ (19%)
HOS Sports-NL	$N = 6$ (4%)	$N = 0$ (0%)	$N = 2$ (2%)	$N = 8$ (7%)

Table VI. SDC and MCID calculations for the HOS ADL-NL and HOS Sports-NL

	SEM	*SDC*	*MCID*
HOS ADL-NL	7.54	21	14
HOS Sports-NL	10.34	29	11

methodology (distribution-based and anchor-based methods), differences in patient cohort and follow-up time. Difference in age at baseline, differences in sports/physical activities or differences in baseline PROM scores, value relevant improvements in scores differently [50, 53]. The duration of follow-up can influence the MCID also as highlighted by Nwachukwu who determined different MCIDs at 1-, 2- and 5-year follow-ups. The SDC we determined was 21 and 29 for the HOS ADL-NL and Sport-NL, respectively, which is high. The MCID we determined was smaller than the SDC. It is, however, important to note that, in our study, a change in scores large enough to represent a clinically relevant change could not be safely detected above measurement error.

Another way of defining the success of hip arthroscopy is through the patient acceptable symptom state (PASS) and by substantial clinical benefit (SCB) [52]. Both known as clinically important outcomes values, which all provide important parameters for determining meaningful improvement after surgery. We have only determined the MCID and not the PASS nor SCB, which might have added more evidence for clinical improvement after surgery in our study.

Other limitations must be mentioned. Our cohort has some heterogeneity regarding the level of activity in patients preoperatively and in surgical procedures performed in patients. Also, we stated that no large cultural adaptation was assumed, considering no large differences in Dutch and American culture. This is an assumption, and differences in patient population due to cultural difference might be present and, therefore, also might slightly influence the differences in outcomes of this study compared to other studies. Only 111 out of 135 included patients could be analyzed. It has been described that <5% loss to follow-up could already lead to small bias [54]. We think this is due to the large number of questionnaires patients were asked to fill in, with overlap in the type of questions. Many patients commented on this. Despite considerable effort to contact all patients, the use of an electronic database instead of papers and to help patients with the questionnaires, loss to follow-up could not be prevented entirely.

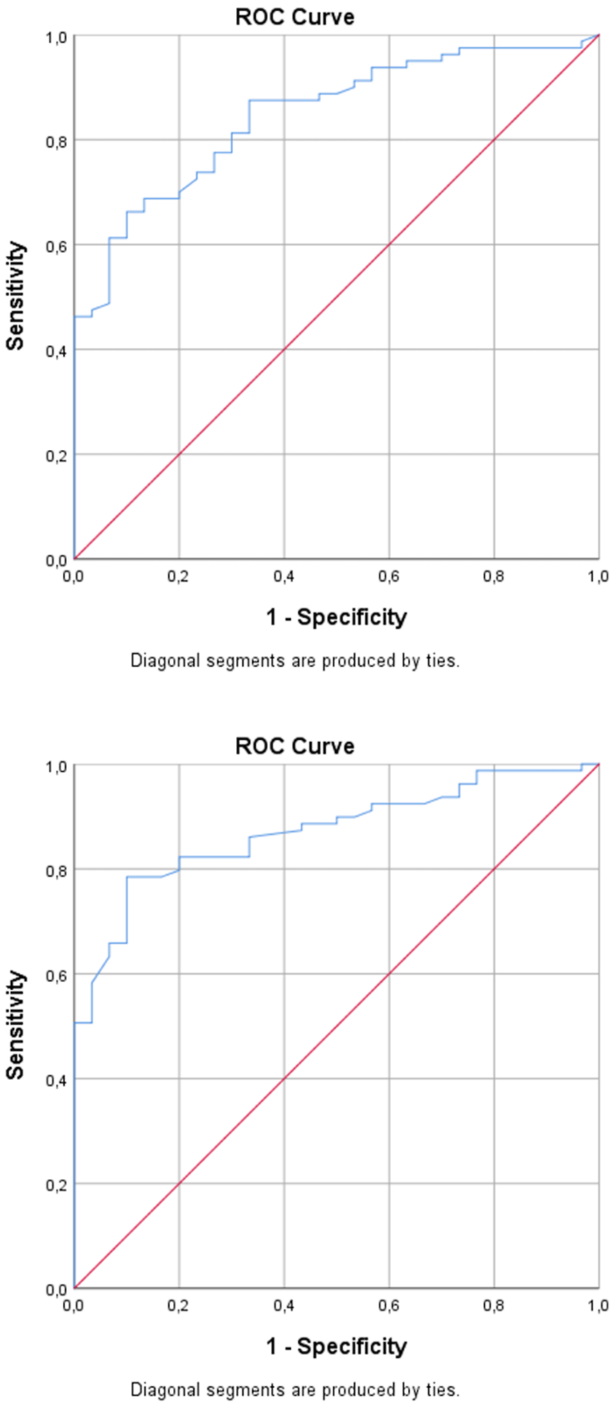

Fig. 2. Receiver operating characteristic curves HOS ADL-NL and HOS Sports-NL.

CONCLUSION

The HOS-NL is a reliable and valid PROM for measuring physical function and outcomes in active and young patients with FAIS.

DATA AVAILABILTY

The data underlying this article will be shared on reasonable request to the corresponding author.

ACKNOWLEDGEMENTS

We would like to express our gratitude to Nicole de Esch, research nurse, who has assisted in informing and including patients, and to Rianne Oomen, research assistant, who has kept track of patients and questionnaires.

FUNDING

None declared.

CONFLICT OF INTEREST STATEMENT

None declared.

REFERENCES

1. 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.
2. Siljander MP, McQuivey KS, Fahs AM *et al.* Current trends in patient-reported outcome measures in total joint arthroplasty: a study of 4 major orthopaedic journals. *J Arthroplasty* 2018; **33**: 3416–21.
3. Lovelock TM, Broughton NS, Williams CM. The popularity of outcome measures for hip and knee arthroplasties. *J Arthroplasty* 2018; **33**: 273–6.
4. Picavet HS, Schouten JS. Musculoskeletal pain in the Netherlands: prevalences and risk groups, the DMC3-study. *Pain* 2003; **102**: 167–78.
5. Röling MA, Pilot P, Krekel PR *et al.* Femoroacetabular impingement: frequently missed in patients with chronic groin pain. *Ned Tijdschr Geneesk* 2012; **156**: A4898.
6. Dwyer T, Whelan D, Shah PS *et al.* Operative versus nonoperative treatment of femoroacetabular impingement syndrome: a meta-analysis of short-term outcomes. *Arthroscopy* 2020; **36**: 263–73.
7. Winge S, Winge S, Kraemer O *et al.* Arthroscopic treatment for femoroacetabular impingement syndrome (FAIS) in adolescens – 5-year follow-up. *J Hip Preserv Surg* 2021; hnab051.
8. Simpson J, Sadri H, Villar R. Hip arthroscopy technique and complications. *Orthop Traumatol Surg Res* 2010; **96**: s68–76.
9. Christensen JC, Marland JD, Miller CJ *et al.* Trajectory of clinical outcomes following hip arthroscopy in female subgroup populations. *J Hip Preserv Surg* **6**; 2019: P 25–32.
10. Byrd JW, Jones KS. Prospective analysis of hip arthroscopy with 10-year follow-up. *Clin Orthop Relat Res* 2010; **468**: 741–6.
11. Kemp JL, Colllins NJ, Makdissi M *et al.* Hip arthroscopy for intra-articular pathology: a systematic review of outcomes with and without femoral osteoplasty. *Br J Sports Med* 2012; **46**: 632–43.
12. MacFarlane RJ, Konan S, El-Huseinny M *et al.* A review of outcomes of the surgical management of femoroacetabular impingement. *Ann R Coll Surg Engl* 2014; **96**: 331–8.
13. Aprato A, Jayasekera N, Villar RN. Does the modified Harris hip score reflect patient satisfaction after hip arthroscopy? *Am J Sports Med* 2012; **40**: 2557–60.
14. 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**: 279–86.
15. 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**: 1185–96.
16. Martin RL, Kelly BT, Philippon MJ. Evidence of validity for the hip outcome score. *Arthroscopy* 2006; **22**: 1304–11.
17. Martin RL, Philippon MJ. Evidence of reliability and responsiveness for the hip outcome score. *Arthroscopy* 2008; **24**: 676–82.
18. Ramisetty N, Kwon Y, Mohtadi N. Patient-reported outcome measures for hip preservation surgery—a systematic review of the literature. *J Hip Preserv Surg* 2015; **2**: 15–27.

19. Kemp JL, Collins NJ, Roos EM *et al.* Psychometric properties of patient-reported outcome measures for hip arthroscopic surgery. *Am J Sports Med* 2013; **41**: 2065–73.
20. Kluzek S, Dean B, Wartolowska KA. Patient-reported outcome measures (PROMs) as proof of treatment efficacy. *BMJ Evidence-Based Med* Published Online First: 04 June 2021. 10.1136/bmjebm-2020-111573.
21. Seijas R, Sallent A, Ruiz-Ibán MA *et al.* Validation of the Spanish version of the Hip Outcome Score: a multicenter study. *Health Qual Life Outcomes* 2014; **13**: 70.
22. Lee YK, Ha YC, Martin RRL *et al.* Transcultural adaptation of the Korean version of the Hip Outcome Score. *Knee Surg Sports Traumatol Arthrosc* 2015; **23**: 3426–31.
23. Costa RMP, Cardinot TM, Carreras Del Castillo Mathias LN *et al.* Validation of the Brazilian version of the hip outcome score (HOS) questionnaire. *Adv Rheumatol* 2018; **58**: 4.
24. Naal FD, Impellizzeri FM, Miozzari HH *et al.* The German Hip Outcome Score: validation in patients undergoing surgical treatment for femoroacetabular impingement. *Arthroscopy* 2011; **27**: 339–45.
25. Mokkink LB, Terwee CB, Knol DL *et al.* The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC Med Res Methodol* 2010; **10**: 22.
26. Mokking LB, Terwee CB, Patrick DL *et al.* The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010; **63**: 737–45.
27. Streiner DL, Norman GR. Mine is bigger than yours: measures of effect size in research. *Chest* 2012; **141**: 595–8.
28. Collins GS, Ogundimu EO, Altman DG. Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. *Stat Med* 2016; **35**: 214–26.
29. Palazón-Bru A, Folgado-de la Rosa DM, Cortés-Castell E *et al.* Sample size calculation to externally validate scoring systems based on logistic regression models. *PloS One* 2017; **12**: e0176726.
30. Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *J Clin Epidemiol* 1993; **46**: 1417–32.
31. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 2016; **15**: 155–63.
32. Tavakol M, Dennick R. Making sense of Cronbach's alpha. *Int J Med Educ* 2011; **2**: 53–5.
33. Taber KS. The use of Cronbach's alpha when developing and reporting research instruments in science education. *Res Sci Educ* 2011; **48**: 1272–96.
34. Naal FD, Impellizzeri FM, Von Eisenhart-Rothe R *et al.* Reproducibility, validity, and responsiveness of the hip outcome score in patients with end-stage hip osteoarthritis. *Arthritis Care Res (Hoboken)* 2012; **64**: 1770–5.
35. Brans E, de Graaf JS, Munzebrock AV *et al.* Cross-cultural adaptation and validation of the Dutch version of the hip and groin outcome score (HAGOS-NL). *PLoS One* 2016; **11**: e0148119.
36. Harris WH. Traumatic arthritis of the hip after dislocation and acetabular fractures: treatment by mold arthroplasty. An end-result study using a new method of result evaluation. *J Bone Joint Surg Am* 1969; **51**: 737–55.
37. Mohtadi NG, Griffin DR, Pedersen ME *et al.* The Development and validation of a self-administered quality-of-life outcome measure for young, active patients with symptomatic hip disease: the International Hip Outcome Tool (iHOT-33). *Arthroscopy* 2012; **28**: 595–605.
38. Griffin DR, Parsons N, Mohtadi NG *et al.* A short version of the International Hip Outcome Tool (iHOT-12) for use in routine clinical practice. *Arthroscopy* 2012; **28**: 611–6.
39. 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**: 1026–34.
40. Alghadir AH, Anwer S, Iqbal A *et al.* Test-retest reliability validity and minimum detectable change of visual analog, numerical rating, and verbal rating scales for measurement of osteoarthritic knee pain. *J Pain Res* 2018; **11**: 851–6.
41. Cohen J. *Statistical Power Analysis for the Behavioral Sciences.* Mahwah, NJ: Lawrence Erlbaum, 1988.
42. Streiner DL, Norman GR. *Health Measurement Scales: A Practical Guide to Their Development and Use.* Oxford: Oxford University Press, 1995.
43. Terwee CB, Bot SD, de Boer MR. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007; **60**: 34–42.
44. Castor EDC. Castor electronic data capture. [online]. 2019. Available at: https://castoredc.com.
45. R Core team. A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2018. Available at: https://www.R-project.org/.
46. Revelle W. Psych: procedures for personality and psychological research, Northwestern University, Evanston, Illinois, USA. 2017. Available at: https://CRAN.R-project.org/package=psychVersion=1.7.8.
47. de Vet HCW, Terwee CB, Knol KL *et al.* When to use agreement versus reliability measures. *J Clin Epidemiol* 2006; **59**: 1033–9.
48. Robin X, Turck N, Hainard A *et al.* pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinform* 2011; **12**: 77.
49. Pieadade SR, Hutchinson MR, Maffulli N. Presently PROMs are not tailored for athletes and high-performance sports practitioners: a systematic review. *JISAKOS* 2019; **4**: 248–53.
50. Terwee CB, Roorda LD, Knol DL *et al.* Linking measurement error to minimal important change of patient-reported outcomes. *J Clin Epidemiol* 2009; **62**: 1062–7.
51. Nwachukwu BU, Beck EC, Kunze KN *et al.* Defining the clinically meaningful outcomes for arthroscopic treatment of femoroacetabular impingement syndrome at minimum 5-year follow-up. *Am J Sports Med* 2020; **48**: 901–7.
52. Ueland TE, Disantis A, Carreira DS *et al.* Patient-reported outcome measures and clinically important outcome values in hip arthroscopy: a systematic review. *JBJS Rev* 2021; **9**: e20.00084.
53. Lauridsen HH, Hartvigsen J, Manniche C *et al.* Responsiveness and minimal clinically important difference for pain and disability instruments in low back pain patients. *BMC Musculoskelet Disord* 2006; **7**: 82.
54. Dettori JR. Loss to follow-up. *Evid Based Spine Care J* 2011; **2**: 7–10.