HIP

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

Roland Thomeé · Pall Jónasson · Kristian Thorborg ·
Mikael Sansone · Mattias Ahldén · Christoffer Thomeé ·
Jon Karlsson · Adad Baranto

Abstract

Purpose There is a lack of standardised outcome measures in Swedish for active, young and middle-aged patients with hip and groin disability. The purpose of this study was to adapt the Danish version of the Copenhagen Hip and Groin Outcome Score (HAGOS) patient-reported outcome instrument for use in Swedish patients and evaluate the adaptation according to the Consensus-Based Standards for the Selection of Health Status Measurement Instruments checklist.

Methods Cross-cultural adaptation was performed in several steps, including translation, back-translation, expert review and pretesting. The final version was evaluated for reliability, validity and responsiveness. Five hundred and two patients (337 men and 167 women, mean age 37, range 15–75) were included in the study.

Results Cronbach's alpha for the six HAGOS-S subscales ranged from 0.77 to 0.89. Significant correlations were obtained with the international Hip Outcome Tool average score ($r_s = 0.37$–0.68; $p < 0.01$) and a standardised instrument, the EuroQol, EQ-5D total score ($r_s = 0.40$–0.60, $p = 0.01$), for use as a measurement of health outcome. Test–retest reliability (intraclass correlation coefficient) ranged from 0.81 to 0.87 for the six HAGOS-S subscales. The smallest detectable change ranged from 7.8 to 16.1 at individual level and 1.6–3.2 at group level. Factor analysis revealed that the six HAGOS-S subscales had one strong factor per subscale. Effect sizes were generally medium or large.

Conclusion The Swedish version of the HAGOS is a valid, reliable and responsive instrument that can be used both for research and in the clinical setting at individual and group level.

Level of evidence Diagnostic study, Level I.

Keywords iHOT12 · EQ-5D · Hip · Groin · COSMIN · Femoro-acetabular impingement · Validity · Reliability · Responsiveness

R. Thomeé (✉)
Section of Physiotherapy, Department of Clinical Neuroscience and Rehabilitation, Institute of Neuroscience and Physiology, Sahlgrenska Academy, Gothenburg University, Göteborg, Sweden
e-mail: roland.thomee@orthop.gu.se

P. Jónasson · M. Sansone · M. Ahldén · C. Thomeé · J. Karlsson · A. Baranto
Department of Orthopaedics, Sahlgrenska University Hospital, Göteborg, Sweden

K. Thorborg
Arthroscopic Centre Amager, Copenhagen University Hospital, Hvidovre, Denmark

Introduction

Until recently, there has been a lack of standardised patient-reported outcome measures for young, active patients with hip and groin disability [16]. The Copenhagen Hip and Groin Outcome Score (HAGOS) was published in 2011 [15], and the international Hip Outcome Tool (iHOT12) was published in 2012 [4]. The iHOT12 has been cross-culturally adapted and validated into a Swedish version, the iHOT12-S [7]. The HAGOS is based on and designed in a manner similar to the KOOS scale and comprises 37 items in six subscales; symptoms (7 items), pain (10 items), function in daily living (5 items), function

in sport and recreation (8 items), participation in physical activities (2 items) and hip- and/or groin-related quality of life (5 items).

Health-related patient-reported outcomes (HR-PROs) are widely used to evaluate the effectiveness of treatment or to compare different interventions in clinical trials. They are questionnaires completed by patients to measure perceptions of their general health or their health in relation to a specific illness or condition. Before an HR-PRO can be used for research or in a clinical setting, it must be standardised, validated and tested for reliability [3]. In 2010, the Consensus-Based Standards for the Selection of Health Status Measurement Instruments (COSMIN) published a checklist which could be used to develop and evaluate HR-PROs [13, 15]. The checklist is designed to be used as a guide in the development of HR-PROs and to evaluate the quality of studies measuring the properties of HR-PROs.

The purpose of this study was cross-culturally to adapt and validate the Swedish HAGOS version, in accordance with the COSMIN checklist.

Materials and methods

The adaptation of the HAGOS to Swedish was performed in several steps, as proposed by Beaton et al. [1].

The original version was translated into Swedish by three of the authors (two orthopaedic surgeons and one physiotherapist) who are fluent in Swedish, well acquainted with the Danish language and experienced in working with patients with hip and groin disability. The three translations were then synthesised into a Swedish version by an expert panel of three orthopaedic surgeons and one physiotherapist. The synthesised version, the result of consensus among the panel, was back-translated into Danish by a native Danish-speaking person, and the translation was subsequently compared with the original version by the same panel. Minor differences between the original and back-translated versions were resolved by consensus among the panel.

A pilot test to check the acceptability of the synthesised version was performed on 10 healthy individuals without any history of hip or groin problems. They were encouraged to make comments with their answers. This was done to ensure that the questions would not be experienced as obtrusive and that non-health care professionals could understand the questions. After the pilot test, minor modifications were made to the synthesised translation, according to consensus among the panel, which mainly involved replacing professional words with more lay terms. Face validity, the degree to which the instrument looks as though it adequately reflects the measured construct [9],

was deemed acceptable according to consensus among the expert panel.

The reliability, validity and responsiveness of the final version, the HAGOS-S, were assessed according to the COSMIN checklist [10] in a clinical study. Five hundred and two patients requiring hip arthroscopy for femoro-acetabular impingement (FAI) based on radiological and clinical criteria completed the questionnaire on their first visit to an experienced hip surgeon. Only those patients requiring hip arthroscopy for FAI were included. At the time of the study, 360 patients (92 % response rate) had completed the questionnaires at 4 months post-operatively. A group of 26 patients completed the HAGOS-S pre-operatively on two separate occasions within 3 weeks for test–retest reliability.

All the patients evaluated their overall hip function on a global perceived effect (GPE) visual analogue scale (VAS) from 0 (extremely poor hip function) to 100 (perfect hip function). A change of 20 points or more on the GPE scale was regarded as representing a clinically relevant change in patient symptoms [3, 5, 10]. Twenty-six patients were included in the test–retest reliability evaluation. To be included in the test–retest evaluation, the patients' condition had to be regarded as clinically stable during this period. It was therefore decided a priori that only patients with a change of fewer than 20 points between test and retest on the VAS could be included in this analysis.

The patients completed the Swedish versions of the EQ-5D [5] and the iHOT12-S [7] to be correlated with the HAGOS-S for construct validity. Their physical activity level was assessed with the Hip Sports Activity Scale (HSAS) [11]. The patients were also asked to use the HSAS to estimate their physical activity level when they were teenagers and before their symptom debut.

The study was approved by the Regional Ethical Review Board, Gothenburg, Sweden, ID: 472-10. All patients gave their informed consent.

Statistical analysis

Statistical analysis was performed using the Statistical Package for the Social Sciences (SPSS) version 21. Most data were ordinal, so nonparametric statistics were used. The level of significance was set at $p < 0.05$. The questionnaires were web based, leaving the patients no option but to answer all the questions. As a result, no individual items were missing.

Reliability

The reliability of an HR-PRO is the degree to which it is free from measurement error [9]. To evaluate the reliability

of an HR-PRO, its internal consistency, test–retest reliability and measurement error must be assessed.

Internal consistency is the degree of interrelatedness between the items [9]. Internal consistency was measured for the six subscales of the HAGOS-S from the baseline values and was deemed good if Cronbach's alpha was between 0.70 and 0.95 [14].

Test–retest reliability is defined as the proportion of the total variance in the measurements which is due to true differences between patients [9]. The intraclass correlation coefficient (ICC), (3.1 two-way mixed effects model absolute agreement) was calculated for each of the six HAGOS-S subscales. An ICC of >0.70 was deemed acceptable [14]. A Wilcoxon's paired test was performed to assess whether there were significant differences in scores between the test occasions.

Measurement error is the systematic and random error of the score, not attributed to the construct that is being measured [9]. Measurement error was expressed as the standard error of the mean (SEM) using the formula $SD \times \sqrt{1-ICC}$, with SD as the standard deviation of scores from all patients at baseline [17]. The smallest detectable change (SDC), a change in a score that exceeds the measurement error, was calculated at individual level as $SEM \times 1.96 \times \sqrt{2}$ and at group level as $SEM \times 1.96 \times \sqrt{2}/\sqrt{n}$ [2].

Validity

Construct validity is the degree to which the scores of a PRO instrument are consistent with a priori hypotheses, based on the assumption that the instrument validly measures the construct that is going to be measured [9]. A principal component factor analysis with varimax rotation and the eigenvalue set at >1 was performed to assess the structural validity of each of the six HAGOS-S subscales. The factor analysis presents the eigenvalue and the variance explained in per cent to indicate the relative strength of the factor. Hypothesis testing was performed using Spearman's correlation coefficient for nonparametric data, comparing the scores from the HAGOS-S with the EQ-5D-S and iHOT12-S scores.

A priori hypotheses were formulated. With the HAGOS and iHOT12 developed for similar patient groups and measuring essentially the same constructs, we expected high correlations (Spearman $r > 0.50$) between the six HAGOS-S subscales and the iHOT12-S average score. We expected a moderate correlation (Spearman $r > 0.30$) between the subscales of the HAGOS-S and the subscales of the EQ-5D-S, but a higher correlation was expected between the HAGOS-S and the mobility, usual activities and pain/discomfort subscales of the EQ-5D-S than with the self-care and anxiety/depression subscales.

Responsiveness

The responsiveness of a PRO instrument is its ability to detect change over time [9]—in the present study, between pre-operatively and a 4-month follow-up. Responsiveness was assessed using Spearman's correlation coefficient, standardised response mean (SRM) and effect size (ES). Correlations between the GPE and the six subscales of the HAGOS-S were measured. The SRM was calculated as the mean change in score divided by the SD of the change. The ES was calculated as the mean change in score divided by the SD of the baseline score [13]. The patients were divided into three groups: those reporting worsening of hip function between pre-operatively and the 4-month follow-up (at least 20 points lower GPE score), those that reported no change in function (0–19 points higher or lower GPE score) and those that reported improved function (at least 20 points higher GPE score). A priori hypotheses were formulated for responsiveness. We hypothesised that the change in the score on the HAGOS-S subscales would correlate with the GPE score with a Spearman correlation coefficient of >0.3. We furthermore hypothesised that the SRM and ES would be higher for those reporting improved hip function between pre-operatively and the 4-month follow-up (at least 20 points lower GPE score) and lower for those reporting worsening of hip function between pre-operatively and the 4-month follow-up (at least 20 points lower GPE score).

Interpretability

Interpretability is defined as the degree to which it is possible to assign qualitative meaning to an instrument's quantitative scores or change in scores [9]. It includes the distribution of total scores and change in scores, floor and ceiling effects and an estimation of the minimal important change (MIC) and/or minimal important difference (MID). Floor and ceiling effects were defined as being present if more than 15 % of patients reported lowest (0) or highest (100) possible scores [15]. The MIC was calculated as $0.5 \times SD$ both at baseline and at 4 months [12].

Results

Baseline characteristics are presented in Table 1. A total of 502 patients completed the HAGOS-S questionnaire at baseline. At the time of the study, 391 patients had reached 4 months post-surgery and 360 (92 %) were available for follow-up. Twenty-six patients completed the questionnaire pre-operatively on two separate occasions with a mean interval of 14 (range 9–20) days (SD 3.3).

Table 1 Baseline characteristics

	Total (n = 502)	Men (n = 337)	Women (n = 165)
Age (years)			
Mean (SD)	37.3 (13.4)	36.0 (12.8)	39.3 (13.9)
Range	15–75	16–71	15–75
Sports activity level HSAS score			
At baseline (n = 459)			
Mean (SD)	3.0 (2.3)	3.4 (2.4)	2.1 (1.9)
HSAS = 0			
n (%)	73 (15.9)	39 (12.9)	34 (21.8)
HSAS = 1–4			
n (%)	270 (58.8)	166 (54.8)	104 (66.7)
HSAS = 5–8			
n (%)	116 (25.3)	98 (32.3)	18 (11.5)
Before symptoms (n = 460)			
Mean (SD)	5.7 (2.0)	6.0 (1.8)	5.0 (2.1)
HSAS = 0			
n (%)	2 (0.4)	1 (0.3)	1 (0.6)
HSAS = 1–4			
n (%)	124 (27.0)	55 (18.2)	69 (44.0)
HSAS = 5–8			
n (%)	334 (72.6)	247 (81.5)	87 (55.4)
As a teenager (n = 460)			
Mean (SD)	6.2 (1.8)	6.6 (1.5)	5.4 (2.2)
HSAS = 0			
n (%)	4 (0.9)	1 (0.3)	3 (1.9)
HSAS = 1–4			
n (%)	70 (15.2)	20 (6.6)	50 (31.9)
HSAS = 5–8			
n (%)	386 (83.9)	282 (93.1)	105 (66.2)
Affected side (n = 502), frequency			
Right			
n (%)	204 (41)	123 (36.5)	81 (49.1)
Left			
n (%)	155 (31)	97 (28.8)	58 (35.2)
Both			
n (%)	143 (28)	117 (34.7)	26 (15.7)
GPE score (n = 495)			
VAS			
Mean (SD)	48.7 (19.9)	48.7 (20.1)	47.5 (19.1)
Range	0–100	0–100	0–90

Reliability

Descriptive statistics and test–retest reliability measurements are presented in Table 2. The ICC ranged from 0.81 to 0.87. No statistically significant difference between the test and retest scores was found. The SDC for the six subscales ranged from 7.8 to 16.1 at individual level and from 1.5 to 3.2 at group level.

The internal consistency for the six subscales ranged from a Cronbach's alpha of 0.77–0.89 (Table 3).

Validity

An exploratory factor analysis of each of the six subscales separately revealed that all the subscales loaded with one strong factor with an eigenvalue over 1.0 explained a large degree of the variance (Table 3).

For the evaluation of the HAGOS-S construct validity, Spearman's correlation coefficients were calculated between the HAGOS-S and EQ-5D-S and the HAGOS-S and iHOT12-S, respectively (Tables 4, 5). All six subscales of the HAGOS-S showed significant correlations with all questions and the total of the iHOT12-S, the EQ-5D-S total score and the EQ-5D-S VAS score.

Responsiveness

Spearman's correlation coefficient between the score on the six HAGOS-S subscales and the GPE scale ranged from 0.40 to 0.62, indicating moderate correlations. The results of the SRM and ES calculations and GPE correlations are presented in Table 6. As hypothesised, the ES and SRM were lower for those reporting a worsening of hip function and higher for those reporting improved hip function at 4 months.

Interpretability

Floor and ceiling effects, present if more than 15 % of the patients reported highest or lowest scores on an individual item, were not found. The distribution of the scores at baseline, at 4 months and the MIC, is presented in Table 7.

Discussion

The principal findings in the present study were that the HAGOS-S is a valid, reliable and responsive HR-PRO, for patients with femoro-acetabular impingement, undergoing hip arthroscopy.

During translation and adaptation, the authors carefully followed a standardised process described in the literature. This should make the adapted version highly comparable with the original version. During the evaluation of the adapted version, the authors carefully followed the COSMIN checklist to ensure the assessment of every psychometric property.

With the development of the COSMIN checklist, health care specialists have a standardised instrument to evaluate the quality of studies measuring PRO instrument properties. The authors have used the COSMIN checklist during

Table 2 Descriptive statistics and test–retest reliability of the six HAGOS-S subscales ($n = 26$)

Sub-scales	Test mean (SD)	Retest mean (SD)	p value	Difference test–retest, mean (SD)	SEM	ICC (95 % CI)	SDC ind	SDC grp
Sympt	56.5 (15.4)	52.7 (17.5)	0.10, ns	7.3 (7.9)	3.0	0.81 (0.64–0.96)	8.3	1.6
Pain	53.4 (14.9)	51.9 (15.9)	0.27, ns	7.6 (4.9)	2.9	0.83 (0.66–0.92)	8.0	1.6
Func DL	61.7 (24.8)	58.7 (22.1)	0.30, ns	8.5 (9.1)	4.9	0.87 (0.72–0.94)	13.6	2.7
Func Sp	41.6 (22.9)	38.0 (23.6)	0.15, ns	11.5 (9.2)	4.5	0.81 (0.62–0.91)	12.5	2.4
Phys	32.2 (29.4)	28.8 (29.5)	0.40, ns	11.1 (14.7)	5.8	0.81 (0.62–0.91)	16.1	3.2
QoL	30.0 (14.1)	29.4 (12.5)	0.77, ns	5.6 (4.5)	2.8	0.85 (0.70–0.93)	7.8	1.5

Sympt symptoms in hip and/or groin, *Pain* pain in hip and/or groin, *Func DL* function in daily living, *Func Sp* function in sport and recreation, *Phys* Participation in physical activities, *QoL* Hip- and/or groin-related quality of life, *SD* Standard deviation, *p* value for Wilcoxon's paired test, between test and retest, *ns* non-significant, *SEM* Standard error of the mean, *ICC* intraclass correlation coefficient, *95 % CI* 95 % confidence interval, *SDC ind* smallest detectable change at individual level, *SDC grp* smallest detectable change at group level

Table 3 Chronbach's alpha (Cα) for internal consistency ($n = 502$) and factor analysis ($n = 502$) with eigenvalue (EV) and degree of variance explained in per cent (%) for the six HAGOS-S subscales

	C_α	EV	%
Sympt	0.77	3.1	45
Pain	0.88	5.1	51
Func DL	0.89	3.5	70
Func Sp	0.89	4.6	57
Phys	0.84	1.7	86
QoL	0.83	3.0	60

Sympt symptoms in hip and/or groin, *Pain* pain in hip and/or groin, *Func DL* function in daily living, *Func Sp* function in sport and recreation, *Phys* participation in physical activities, *QoL* hip- and/or groin-related quality of life

the design and reporting of the present study. We found the checklist easy to follow, but, as it does not as yet conclude what constitutes adequate measurement qualities, criteria proposed in the literature were used during calculations in the present study.

Study population

The HAGOS was developed for young, active patients with hip disorders, but it has been validated on a population between 18 and 60 years of age. In the present study, we included some 500 patients, some younger and some older (15–75 years), and only patients with FAI. Some floor and ceiling effects were experienced. We believe, however, that the HAGOS-S can also be utilised for older patients and for patients with other hip disorders, but future studies are needed to clarify this.

Reliability

All subscales showed very good homogeneity, with an internal consistency between 0.77 and 0.89, as measured with Cronbach's alpha.

With an ICC between 0.81 and 0.89 for the six subscales, the test–retest reliability of the HAGOS-S was found to be very good and in agreement with the ICC reported in the original publication [15].

Table 4 Spearman's correlation coefficients at baseline for HAGOS-S subscales and EQ-5D-S subscales, total score and VAS ($n = 495$)

	HAGOS-S					
	Sympt	Pain	Func DL	Func Sp	Phys	QoL
Mobility	−0.54**	−0.48**	−0.57**	−0.51**	−0.25**	−0.44**
Self-care	−0.20**	−0.21**	−0.27**	−0.23**	−0.10**	−0.14**
Usual activities	−0.31**	−0.24**	−0.33**	−0.37**	−0.43**	−0.42**
Pain/discomfort	−0.50**	−0.39**	−0.47**	−0.38**	−0.22**	−0.45**
Anxiety/depression	−0.24**	−0.21**	−0.24**	−0.27**	−0.29**	−0.50**
Total score	0.57**	0.47**	0.57**	0.52**	0.40**	0.60**
VAS	0.46**	0.33**	0.44**	0.45**	0.40**	0.55**

Sympt symptoms in hip and/or groin, *Pain* pain in hip and/or groin, *Func DL* function in daily living, *Func Sp* function in sport and recreation, *Phys* participation in physical activities, *QoL* hip- and/or groin-related quality of life

** Significant correlation $p < 0.01$

Table 5 Spearman's correlation coefficients at baseline for the 12 items and average score for the iHOT12-S and the six subscales of the HAGOS-S ($n = 495$)

	HAGOS-S					
	Sympt	Pain	Func DL	Func Sp	Phys	QOL
Q1	0.44**	0.53**	0.50**	0.45**	0.25**	0.47**
Q2	0.52**	0.61**	0.70**	0.54**	0.17**	0.36**
Q3	0.45**	0.59**	0.59**	0.56**	0.38**	0.48**
Q4	0.45**	0.25**	0.24**	0.29**	0.10*	0.26**
Q5	0.42**	0.49**	0.57**	0.56**	0.33**	0.46**
Q6	0.34**	0.29**	0.34**	0.50**	0.30**	0.35**
Q7	0.41**	0.50**	0.47**	0.53**	0.39**	0.45**
Q8	0.35**	0.50**	0.55**	0.47**	0.25**	0.44**
Q9	0.40**	0.45**	0.48**	0.43**	0.27**	0.39**
Q10	0.33**	0.38**	0.35**	0.34**	0.27**	0.54**
Q11	0.27**	0.30**	0.28**	0.37**	0.29**	0.47**
Q12	0.28**	0.29**	0.27**	0.36**	0.33**	0.55**
Total	0.58**	0.65**	0.68**	0.67**	0.37**	0.61**

Sympt symptoms in hip and/or groin, *Pain* pain in hip and/or groin, *Func DL* function in daily living, *Func Sp* function in sport and recreation, *Phys* participation in physical activities, *QoL* hip- and/or groin-related quality of life

* Significant correlations $p < 0.05$

** Significant correlations $p < 0.01$

In order to express the patients' clinical change in hip status, it was decided in the present study to use a VAS to determine whether significant changes in patient symptoms had occurred. A change of 20 mm or more was considered clinically important. Minimal important changes on a pain VAS have been found to range from 13 to 30 mm [3, 5, 10, 12].

The SDC for the six subscales of the HAGOS-S at individual level was at a clinically acceptable level (between 7.8 and 16.1), and the HAGOS-S could therefore

be recommended for use in individual patients. A change of 20 points as used in this study for a clinically relevant change in GPE can thus also be recommended as a clinically relevant change at individual level in the HAGOS-S. The low SDC values at group level (between 1.5 and 2.7) strongly indicate that the HAGOS-S is very useful for group comparisons.

Validity

Significant correlations were found between the HAGOS-S subscales and the EQ-5D-S total score, ranging from $r_s = 0.40$ to $r_s = 0.60$. Significant correlations were also found between the HAGOS-S and the iHOT12-S average score, ranging from $r_s = 0.37$ to $r_s = 0.68$, which was as hypothesised, apart from the HAGOS-S subscale of physical activity. Significant correlations were found between the HAGOS-S subscales and EQ-5D-S subscales, ranging from $r_s = -0.10$ to $r_s = -0.57$. As hypothesised, somewhat lower correlations were found for the EQ-5D-S subscales of self-care (average $r_s = -0.19$) and anxiety/depression (average $r_s = -0.29$) compared with the subscales of mobility, usual activities and pain/discomfort (average $r_s = -0.47$, -0.35, -0.40, respectively). The latter three subscales thus correlated more highly with the HAGOS-S than hypothesised.

The factor analysis revealed that the six HAGOS-S subscales had one strong factor per subscale, which is in accordance with the original HAGOS [15].

Responsiveness

The GPE score correlated strongly with the HAGOS-S subscales, ranging from $r_s = 0.40$ to $r_s = 0.68$. As hypothesised, the SRM and ES were lower for patients reporting little clinical change in hip status and higher for

Table 6 Responsiveness of the HAGOS-S measured against different change scores on the GPE scale

HAGOS-S	GPE ($n = 495$)	GPE 20 p lower at 4 months ($n = 26$)		GPE within $\pm$ 20 p at 4 months ($n = 147$)		GPE 20 p higher at 4 months ($n = 187$)	
	r_s	SRM	ES	SRM	ES	SRM	ES
Sympt	0.52**	−0.24	−0.19	0.51	0.48	1.42	1.37
Pain	0.62**	−0.27	−0.22	0.63	0.54	1.61	1.42
Func DL	0.58**	−0.39	−0.23	0.47	0.37	1.31	1.17
Func Sp	0.58**	−0.59	−0.36	0.33	0.35	1.45	1.55
Phys Act	0.40**	−0.62	−0.40	0.20	0.23	0.74	1.07
QoL	0.59**	−0.56	−0.44	0.46	0.48	1.38	1.87

Sympt symptoms in hip and/or groin, *Pain* pain in hip and/or groin, *Func DL* function in daily living, *Func Sp* function in sport and recreation, *Phys* participation in physical activities, *QoL* hip- and/or groin-related quality of life, *GPE* global perceived effect, *SRM* standardised response mean, *ES* Effect size

** Significant correlations $p < 0.01$

Table 7 The HAGOS-S score at baseline and at 4 months with frequencies of lowest (floor effect) and highest (ceiling effect) scores

	Mean	SD	Median	Min–max	Floor effect (%)	Ceiling effect (%)	MIC
Baseline (n = 502)							
Sympt	50.1	18.5	50.0	0–100	1 (0.2)	3 (0.6)	9.3
Pain	54.9	19.3	52.5	0–100	1 (0.2)	1 (0.2)	9.7
ADL	57.5	23.5	55.0	0–100	3 (0.6)	20 (4.0)	11.8
Sport	37.4	21.6	34.4	0–100	13 (2.6)	1 (0.2)	10.8
Phys Act	25.7	26.1	25.0	0–100	158 (31.5)	12 (2.4)	13.1
Qol	30.0	17.6	30.0	0–100	13 (2.6)	1 (0.2)	8.8
Four months (n = 360)							
Sympt	67.2	18.9	67.9	14–100	0 (0.0)	7 (1.9)	9.5
Pain	73.0	19.5	75.0	10–100	0 (0.0)	25 (6.9)	9.8
ADL	75.3	22.0	80.0	15–100	0 (0.0)	61 (16.9)	11.0
Sport	56.6	26.1	56.3	0–100	2 (0.6)	14 (3.9)	13.1
Phys Act	41.9	33.8	37.5	0–100	79 (21.9)	37 (10.3)	16.9
Qol	49.3	25.3	50.0	0–100	5 (1.4)	16 (4.4)	12.7

Sympt symptoms in hip and/or groin, *Pain* pain in hip and/or groin, *ADL* function in daily living, *Sport* function in sport and recreation, *Phys Act* participation in physical activities, *QoL* hip- and/or groin-related quality of life, *SD* standard deviation, *MIC* minimal important change

patients reporting a larger clinical change in hip status, indicating good responsiveness of the HAGOS-S. Clinically, most of the patients had recovered well (although not completely) after 4 months. Larger ES and SRM can thus be expected at 4 months compared with 12 months, for example.

Interpretability

Floor and ceiling effects were detected in the HAGOS-S. At baseline, 31.5 % of the patients obtained the lowest score on the subscale of *participation in physical activities*. The two questions in the subscale ask: *Are you able to participate in your preferred physical activities for as long as you would like?* and *Are you able to participate in your preferred physical activities at your normal performance level?* with the alternatives: Always—Often—Sometimes—Rarely—Never. It is not surprising that many patients with hip and/or groin disability choose the alternative *never*. At 4 months, however, fewer patients (21.9 %) chose the alternative *never*. Future studies will show whether this apparent floor effect is present in the long term. At 4 months, there is a ceiling effect (16.9 %) in the *function in daily living* subscale, indicating that the sensitivity of this subscale can be limited in this patient population.

Taken together, the HAGOS-S with its six subscales can be recommended for measuring both improvement and deterioration over time in the study population.

When developing the COSMIN checklist, no consensus was reached about the method that should be used to measure the MIC [10]. The MIC is supposed to measure

the minimal change in score that the patient regards as important. The rule of thumb that the MIC can be estimated as half an SD was proposed by Norman et al. [12], and, as long as no consensus is reached on the methods by which the MIC should be measured, the authors find this simple rule as good as any other. Applying this rule to the data gave an MIC of 9–17 for the HAGOS-S subscales at baseline and the 4-month follow-up. In the present study, the SDC at individual level is slightly higher than the MIC for some of the HAGOS subscales and slightly lower for other subscales at individual level both at baseline and at the 4-month follow-up. Results at individual level should therefore be interpreted with caution.

Each of the six HAGOS subscales can be used independently to identify changes in certain aspects of patients' symptoms and for certain subpopulations at both group and individual level.

The present data are in agreement to a very large extent, in terms of reliability, validity and responsiveness, with the original study [15] of patients with hip and groin disability. Kemp et al. [8] recently evaluated, on a small subpopulation of 50 patients undergoing hip arthroscopic surgery, the reliability, validity, responsiveness and interpretability of five HR-PROs [Copenhagen Hip and Groin Outcome Score (HAGOS), Hip Disability and Osteoarthritis Outcome Score (HOOS), Hip Outcome Score (HOS), International Hip Outcome Tool (iHOT-33) and modified Harris hip score (MHHS)]. They concluded that some of the psychometric properties of the HAGOS were reduced, based on the fact that the HAGOS subscale related to activities of daily living showed a ceiling effect, which is in agreement with the present study. It is, however, not surprising that, as

patients get better, they become symptom free in activities of daily living before they are symptom free in more sport-related activities. In a recent study, Hinman et al. [6] searched for the best HR-PRO for 30 patients with femoro-acetabular impingement in terms of test–retest reliability. They were able to demonstrate that the majority of the questionnaires, including the HAGOS, were reliable and precise enough for use at group level, which is in agreement with the present study, which also showed that the HAGOS can be used at individual level. Taken as a whole, this study shows that the HAGOS is a highly relevant measurement for patients with unspecific hip and groin pain, as well as for patients with femoro-acetabular impingement, undergoing hip arthroscopy.

Conclusion

The HAGOS-S showed good reliability, validity and responsiveness and can be used both for research and clinically at individual and group level in active patients with hip and/or groin pain.

References

1. Beaton DE, Bombardier C, Guillemin F, Ferraz MB (2000) Guidelines for the process of cross-cultural adaptation of self-report measures. Spine 25:3186–3191
2. Busija L, Osborne RH, Nilsdotter A, Buchbinder R, Roos EM (2008) Magnitude and meaningfulness of change in SF-36 scores in four types of orthopedic surgery. Health Qual Life Outcomes 6:55
3. Fitzpatrick R, Davey C, Buxton MJ, Jones DR (1998) Evaluating patient-based outcome measures for use in clinical trials. Health Technol Assess 2(i–iv):1–74
4. Griffin DR, Parsons N, Mohtadi NG, Safran MR (2012) A short version of the International Hip Outcome Tool (iHOT-12) for use in routine clinical practice. Arthroscopy 28:611–616; quiz 616–618
5. Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, Bonsel G, Badia X (2011) Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res 20:1727–1736
6. Hinman RS, Dobson F, Takla A, O'Donnell J, Bennell KL (2013) Which is the most useful patient-reported outcome in femoro-acetabular impingement? Test-retest reliability of six questionnaires. Br J Sports Med
7. Jonasson P, Karlsson J, Baranto A, Swärd L, Sansone M, Thomeé C, Ahldén M, Thomeé R (2013) A standardized outcome measure for pain, symptoms and physical function in patients with femoroacetabular impingement. Cross-cultural adaptation and validation of the iHOT12 according to the COSMIN checklist. Submitted
8. Kemp JL, Collins NJ, Roos EM, Crossley KM (2013) Psychometric properties of patient-reported outcome measures for hip arthroscopic surgery. Am J Sports Med
9. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HC (2009) Cosmin checklist manual
10. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HC (2010) The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res 19:539–549
11. Naal FD, Miozzari HH, Wyss TF, Notzli HP (2011) Surgical hip dislocation for the treatment of femoroacetabular impingement in high-level athletes. Am J Sports Med 39:544–550
12. Norman GR, Sloan JA, Wyrwich KW (2003) Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. Med Care 41:582–592
13. Streiner DL, Norman GR (2008) Health measurement scales: A practical guide to their development and use. Oxford University Press, New York
14. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, Bouter LM, de Vet HC (2007) Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 60:34–42
15. Thorborg K, Holmich P, Christensen R, Petersen J, Roos EM (2011) The Copenhagen Hip and Groin Outcome Score (HAGOS): development and validation according to the COSMIN checklist. Br J Sports Med 45:478–491
16. Thorborg K, Roos EM, Bartels EM, Petersen J, Holmich P (2010) Validity, reliability and responsiveness of patient-reported outcome questionnaires when assessing hip and groin disability: a systematic review. Br J Sports Med 44:1186–1196
17. Weir JP (2005) Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. J Strength Cond Res 19:231–240