


Validation of a German version of the International Hip Outcome Tool (G-iHOT33) according to the COSMIN checklist: how much improvement is clinically relevant?

Florian Baumann¹  · Johannes Weber¹ · Florian Zeman² · Michael Müller¹ · Matthias Lahner³ · Michael Nerlich¹ · Stefan Fickert^{4,5}

Received: 28 April 2015 / Published online: 29 September 2015
© Springer-Verlag Berlin Heidelberg 2015

Abstract

Purpose The evaluation of the subjective health-related quality of life is especially for young, active patients with hip joint disorders important. The MAHORN study group has recently developed the “International Hip Outcome Tool” (iHOT33), a self-administered patient-reported outcome tool, which includes questions on the patient’s symptoms, functional and sports limitations as well as social, emotional, and occupational limitations. The purpose of this study was to adapt and validate a German version of the iHOT33 according to the COSMIN checklist.

Methods To validate the G-iHOT33, we conducted a prospective multicenter cohort study on patients with hip disorders and a score ≥ 4 on the modified Tegner Activity Scale. The patients completed the G-iHOT33 questionnaire twice at intervals of at least 2 weeks. In addition, we

recorded the Hip Outcome Score (HOS), a modified Tegner Activity Scale (TAS), the EuroQol-5D (EQ5-D), and a subjective assessment of the limitations. Evaluation of psychometric properties was conducted following the COSMIN checklist for validation of health status measurement instruments. The methodical testing for reliability included internal consistency, test–retest reliability, and measurement error. For testing of validity, we analyzed construct validity, hypotheses testing, interpretability, and responsiveness.

Results Between December 2013 and December 2014, eighty-three consecutive patients completed both questionnaires and were available for data analysis. Cronbach’s α was 0.97 (95 % CI 0.96, 0.99) confirming internal consistency and test–retest reliability of the G-iHOT-33 was high with an ICC = 0.88 (95 % CI 0.80, 0.99). All a priori hypotheses were confirmed, further, no floor- or ceiling-effects occurred. The G-iHOT33 showed good responsiveness with a minimal important change (MIC) of 10 points.

Conclusions The German translation of the iHOT-33 (G-iHOT-33) is a viable tool for the evaluation of active patients with a hip disorder. Following the complete COSMIN checklist, we could prove that G-iHOT33 is a reliable, valid, and responsive PRO measurement tool. We could show that the minimal important change, a change of health condition the patient discerns is 10 points in the G-iHOT33 scale. This is the first study providing results on psychometric properties of the iHOT33 subscales.

Level of evidence 1b validating cohort study.

Keywords Patient-reported outcome (PRO) · Hip arthroscopy · Femoro-acetabular impingement syndrome (FAI) · Responsiveness · German International Hip Outcome Tool 33 (G-iHOT33) · COSMIN checklist

Electronic supplementary material The online version of this article (doi:10.1007/s00402-015-2336-1) contains supplementary material, which is available to authorized users.

✉ Florian Baumann
florian.baumann@ukr.de

¹ Department of Trauma Surgery, Regensburg University Medical Center, Regensburg, Germany

² Center for Clinical Studies, Regensburg University Medical Center, Regensburg, Germany

³ Department of Orthopedic Surgery, St. Josef-Hospital, Ruhr-University Bochum, Bochum, Germany

⁴ Medical Faculty Mannheim, University Medical Centre Mannheim, Heidelberg, Germany

⁵ Sporthopaedicum Straubing, University of Mannheim, Mannheim, Germany

Introduction

Patient-reported outcome (PRO) measurements have become an important tool to evaluate activities, limitations in everyday life and the quality of life in hip disorders. Conventional questionnaires focus on patients with osteoarthritis or undergoing hip arthroplasty with a limited activity level [1–4]. Over the last decade, the better understanding of specific hip pathologies has evolved joint-preserving procedures such as hip arthroscopy or surgical hip dislocation [5]. The mainly young and active patients undergoing these joint-preserving procedures have different expectations and aims of their surgery; conventional outcome tools do not reflect their situation adequately [6–8]. Recently the Multi-center Arthroscopy of the Hip Outcomes Research Network (MAHORN) study group developed a new PRO questionnaire with special concern for young, active patients with hip disorders [9]. Compared to other questionnaires, the iHOT-33 includes inquiries of limitations in social interactions, emotional issues and working life. The iHOT-33 has shown a high reliability and validity. So far, there has been no German version of the iHOT-33.

Studies evaluating measurement properties have to meet a high methodological quality. The COSMIN checklist (COnsensus-based Standards for the selection of health status Measurement INstruments) is a consensus-based checklist to evaluate the methodological quality of studies on measurement properties of health status measurement instruments based on an international Delphi study in 2010 [10].

The purpose of this study was to validate a German version of the iHOT-33 according to the COSMIN checklist and evaluate the psychometric properties of the iHOT33 subscales.

Materials and methods

International Hip Outcome Tool (iHOT-33)

In 2012, the Multicenter Arthroscopy of the Hip Outcomes Research Network (MAHORN) developed the International Hip Outcome Tool (iHOT-33), a self-administrated questionnaire comprising of 33 items. [9]. The questionnaire consists of four different subscales indicating symptoms/function, sports activity, social, and occupational limitations. The patient is asked to consider the problems of the past month and to indicate the severity on a 100-mm horizontal line by marking it with a slash. Each question has equal weight so that the mean of all questions amounts to the score result ranging from 0 to 100. A score of 100 indicates full function and no symptoms, whereas a score of zero signifies maximum limitations and extreme

symptoms. The iHOT-33 has shown a high internal consistency, construct validity, and responsiveness [8, 9, 11–13].

Adaption of the iHOT-33

The translation of the iHOT-33 into German was carried out following the guidelines of the American Academy of Orthopedic Surgeons (AAOS) Outcomes Committee [14]. According to these guidelines, an informed and an uninformed translator translated the iHOT-33 from English into German independently. After consolidation of both translations, a German linguist reviewed the German version of the questionnaire. Two native speaking translators (informed and uninformed) re-translated this German version into English. This version was verified for consistence. Finally, the German questionnaire was tested for comprehensibility in 20 patients with a hip disorder. The translation process was supervised and documented in a survey report.

Validation study

A prospective multicenter clinical trial was performed to evaluate reliability, validity, and responsiveness of the German version of the iHOT-33. Inclusion criteria were a history of a hip disorder, a score of ≥ 4 on a modified Tegner Activity Scale [15], and sufficient reading and comprehension capacity. Another disorder of the back or the contralateral lower extremity, a score less than 4 on the modified Tegner Activity Scale, a mental disorder, or a lack of informed consent to participation were exclusion criteria [9]. All patients were seen in an outpatient setting. The patients primarily completed the questionnaire before seeing the orthopedist. For evaluation of test–retest reliability, the patients completed a second questionnaire after a minimum of 2 weeks. The patients were asked to answer the questions according to their current status and return the forms by mail. We reminded all patients who did not answer within 6 weeks by telephone. All patients had given their written informed consent to participate in this study. The local ethical committee approved the study in November 2013.

Questionnaire

In addition to the G-iHOT-33, the questionnaire consisted of the following scores.

Hip Outcome Score (HOS)

The HOS is an established 31-item PRO tool to evaluate activities, limitations in everyday life, and quality of life of

patients with a hip disorder. It comprises of two subscales on activity of daily life and sports activities. The patient is asked to answer the questions considering the past week. Scores range 0–100, higher scores represent a better function and a higher level of activity [16]. The HOS has been validated and published in German in 2011 [17].

Modified Tegner Activity Scale (mTAS)

The TAS is a 10-level activity scale reflecting the patient's currently highest level of sports activity or other routine activities. Initially it was designed as a complement to other functional scores of the knee joint and is the most commonly used activity scoring tool [15]. Although there is no validation study of the hip modification of the TAS, it is also well established and widely used [9, 13, 18, 19]. A score greater than 4 was an inclusion criterion for the evolution study of the iHOT-33 by Mohtadi [9]. Hence, we included the scale to our questionnaire to achieve a similar cohort.

EuroQol-5D (EQ5-D)

The EQ5-D is a global quality of life questionnaire consisting of a 5-item assessment of the health status regarding mobility, self-care, usual activities, pain/discomfort, and anxiety/depression [20]. The second part of the EQ5-D consists of a 200-mm analog scale concerning the patient's assessment of the current global health status. The EQ5-D has been adapted to German and is validated for a number of health compromising conditions [21].

Subjective assessment

The patient was asked to assess his current limitations concerning function (pain, ROM, etc.), sport/leisure activities, employment/housekeeping, and social interaction/quality of life. The limitations should be estimated in percent from 0 % = no limitation at all to 100 % = maximum.

The second set of questions also included an evaluation of whether the condition of their hip joint was 'much better', 'somewhat better', 'unchanged', 'somewhat worse', or 'much worse' compared to the primary evaluation.

Statistical analysis

Questionnaires with any missing data or unclear marking were excluded from the analysis. Statistical analysis was performed using the software package SPSS (Version 22, SPSS Inc., Chicago, Illinois). Unless otherwise stated, descriptive data are given as mean \pm standard deviation. The level of significance was defined at $p < 0.05$ for all tests.

Methodological testing according to the COSMIN checklist

Reliability

Reliability is the degree to which the measurement is free from measurement error [22]. To evaluate reliability, internal consistency, test–retest reliability, and measurement error are calculated.

Internal consistency

Internal consistency is described as the degree of interrelatedness among items [22]. Cronbach's alpha was calculated in total and for each subscale separately. Sufficient internal consistency was assumed for a Cronbach's α greater than 0.7 [23].

Test–retest reliability

Test–retest reliability is the extent to which results of the same patient in the same health condition remain unchanged over time [22]. According to the recommendation of the COSMIN manual, the retest was performed after a minimum of 2 weeks after outpatient consultation to avoid recollection of the answers and changes in health condition. Intraclass correlation coefficients (ICC) were calculated for all patients indicating an unchanged condition of their hip joint since the primary evaluation. For an ICC greater than 0.7 sufficient test–retest reliability was assumed [23].

Measurement error

The measurement error is the systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured [22]. The Standard Error of Measurement (SEM) was calculated using the formula $SD/\sqrt{1 - ICC}$ (SD = standard deviation; ICC = intraclass correlation coefficient) [23]. The smallest detectable change (SDC) reflects the smallest individual change in score that can be interpreted as a real change. It was calculated by the $SEM \times 1.96 \times \sqrt{2}/\sqrt{n}$ [23].

Validity

Validity is the degree to which a questionnaire measures the construct it purports to measure [22].

Construct validity

Since there is no gold standard in the measurement of PRO, validity was rated as construct validity. Construct validity

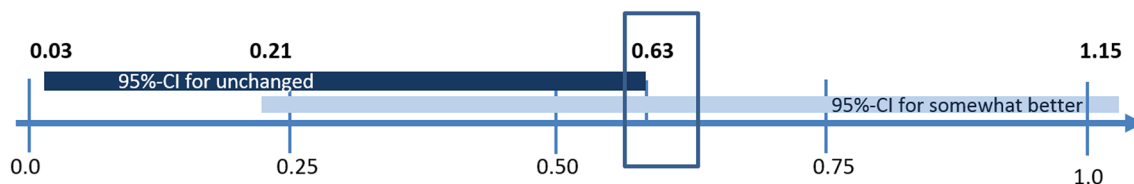


Fig. 1 Estimation of the effect size

is the degree to which the scores of a questionnaire are consistent with questionnaires measuring the same construct. To validate the G-iHOT33, Pearson's correlation coefficient was calculated between the iHOT33 subscales and a subjective rating by the patient as well as between iHOT33 subscales 1 and 2 (symptoms/function and sports activity) and the HOS subscales and between the social/emotional subscale of the iHOT33 (No. 4) to the EQ-5D score. For a correlation coefficient $r < 0.3$, a poor and for $r > 0.7$ a good correlation was assumed.

Hypothesis testing

To analyze construct validity, we tested a priori hypotheses [22]. We hypothesized that the G-iHOT33 would correlate well with the other subjective scales like the HOS and the EQ-5D. Therefore, a Pearson correlation coefficient $r > 0.7$ was expected. Low correlations $r < 0.3$ were expected between the G-iHOT33 and the mTAS. The subjective global rating of change (GRC) was correlated with the mean difference between the G-iHOT33 scores at T2–T1. We hypothesized that the changes in the G-iHOT33 score would correlate with the subjective evaluation of the patient [19, 24].

Responsiveness

Responsiveness is the ability of a questionnaire to detect a change over time in the construct to be measured [22]. According to Terwee et al. [23], responsiveness was demonstrated by comparing the smallest detectable change (SDC) to the minimal important change (MIC). Responsiveness was confirmed if the $SDC < MIC$.

Interpretability

Interpretability is the ability to transform a qualitative effect into a quantitative score [22]. The minimal important change (MIC) was estimated by the dividing the standard deviation (SD) by two as described by Norman et al. [25]. In addition, we used an anchor-based method to evaluate responsiveness. At T2, the patients were asked to rate whether the current condition of their hip joint was 'much better', 'somewhat

better', 'unchanged', 'somewhat worse', or 'much worse' compared to the condition of the primary evaluation. The effect size (ES) was calculated by the mean change of the score/SD. The 95 % CI of the effect size of the "somewhat better" group was compared to the ES of the "unchanged" group to estimate the true MIC (Fig. 1).

Another quality criterion for content validity is the absence of floor and ceiling effects. If more than 15 % of patients score highest (100) or lowest (0) value in the G-iHOT33, extreme outcome values might not be represented adequately [23].

Results

Demographic data and generalizability

Between December 2013 and December 2014 eighty-three patients completed both questionnaires and were available for data analysis. The cohort comprised of 24 women (29 %) and 59 men (71 %). The mean age was 33.7 ± 11.8 years (range 14–63). Demographic data and diagnosis-related score results are provided in Table 1. The second questionnaire was completed on average 28.5 ± 31.7 days (range 14–194) after the first. Missing items were found in 92 of 5146 items in total (1.78 %). Questionnaires containing missing items or unclear marking were excluded from the analysis. Missing items occurred randomly; there was no accumulation of missing items in any unit of the questionnaire.

Reliability

A Cronbach's α of 0.97 (95 % CI 0.96, 0.99) showed excellent internal consistency for the G-iHOT33. Internal consistency was appropriate with a Cronbach's α of between 0.88 and 0.96 for each subscale. The intraclass correlation coefficient (ICC) was 0.88 (95 % CI 0.80, 0.93) for all patients indicating an unchanged condition of their hip joint since their primary evaluation. The overall SEM was 8.9. Hence, the smallest detectable change (SDC) reflecting the smallest individual change in score that can be interpreted as a real change was 2.7.

Table 1 Demographic data and diagnosis-related score results

Diagnosis	Number of patients	Mean age	Gender		iHOT-33	HOS	mTAS
			Male	Female			
FAI	31 (37.3 %)	28.5 (\pm 9.6)	26	5	58.0 (\pm 19.0)	81.6 (\pm 12.6)	5.9
Osteoarthritis	16 (19.3 %)	47.4 (\pm 9.0)	12	4	53.2 (\pm 29.0)	76.9 (\pm 17.0)	4.3
Hip dysplasia	13 (15.7 %)	24.7 (\pm 8.2)	5	8	54.1 (\pm 14.9)	78.9 (\pm 10.2)	5.6
Muscular imbalance	6 (7.2 %)	31.7 (\pm 11.1)	3	3	71.6 (\pm 13.4)	85.2 (\pm 6.9)	5.3
Not specified	17 (20.5 %)	37.4 (\pm 7.7)	13	4	65.6 (\pm 39.3)	82.2 (\pm 23.5)	4.9
Total	83 (100 %)	33.7 (\pm 11.8)	59	24	59.3 (\pm 25.9)	80.9 (\pm 15.7)	5.3

Table 2 Pearson correlation coefficients between functional scores

<i>n</i> = 83	G-iHOT33	HOS	mTAS
HOS	0.86 (<i>p</i> = 0.000)	1	0.02 (<i>p</i> = 0.87)
mTAS	0.07 (<i>p</i> = 0.56)	0.02 (<i>p</i> = 0.87)	1
EQ-5D	0.70 (<i>p</i> = 0.000)	0.63 (<i>p</i> = 0.000)	0.071 (<i>p</i> = 0.52)

Validity

The assessment of the construct validity showed a good correlation between the subscales of G-iHOT-33, HOS, and EQ-5D (Table 2). With a Pearson correlation coefficient of 0.07, there was only a poor correlation between G-iHOT33 and the mTAS. Therefore, all hypotheses could be confirmed. Adequate responsiveness of the G-iHOT33 could be demonstrated with a higher value of MIC (12.7) compared to SDC (2.7). According to the GRC-dependent 95 % CI of the ES, the estimation of ES was confirmed (Fig. 1). In addition, there was a good correlation between the global rating of change (GRC) and the mean difference between the G-iHOT33 scores at T2–T1 (Table 3; Fig. 2). According to the ES of 0.6 the minimal important change, a change that reflects a clinically relevant improvement is approximately ten points on the iHOT33.

There was no floor effect for the G-iHOT33, since there was no patient with a score value of zero. Only five patients (6 %) scored a maximum score of 100, which is why relevant ceiling effects could be declined.

The Pearson correlation coefficients between the subscales concerning symptoms/function, sports activity, social, and occupational limitations and the subjective rating are shown in Table 4. We found a good correlation between the subscales and the HOS subscales and the patient's subjective rating for each of the subscales. The highest correlation coefficient of $r = 0.83$ was found between the social subscale and the EQ-5D. However, none of the subscales showed floor- or ceiling-effects.

Discussion

The present study evaluated the German version of the iHOT33. This German version of the iHOT33 provides sufficient validity, reliability, and responsiveness for the evaluation of physically active patients with non-arthritic hip problems. In addition, we are able to present results on psychometric properties of the iHOT33 subscales on symptoms/function, sports activity, social, and occupational limitations separately. This is the first study validating the 33-item iHOT following the complete COSMIN checklist.

Since femoro-acetabular impingement syndrome (FAI) has been identified as a risk factor for osteoarthritis of the hip joint, the development of joint-preserving procedures has been advanced in the last decade. Due to technical improvements, hip arthroscopy has become a successful procedure to relieve pain and to restore clinical function in FAI [5, 8, 26–29]. Patient-reported outcome tools are

Table 3 Change sensitivity of the iHOT33 based on global rating of change

Global rating of change (GRC)	<i>N</i>	Mean difference	SD	Mean difference 95 % CI	Minimum	Maximum	Effect size (ES)	95 % CI of ES
Somewhat worse	6	−6.54	12.99	−20.17 7.09	−23.76	14.33	−0.50	−1.34 0.37
Unchanged	46	3.95	11.98	0.40 7.51	−11.45	62.27	0.33	0.03 0.63
Somewhat better	22	5.61	8.16	1.99 9.23	−11.67	25.52	0.69	0.21 1.15
Much better	9	36.36	34.15	10.12 62.61	−15.45	92.52	1.06	0.21 1.88

Fig. 2 Mean difference of the iHOT33 for GRC

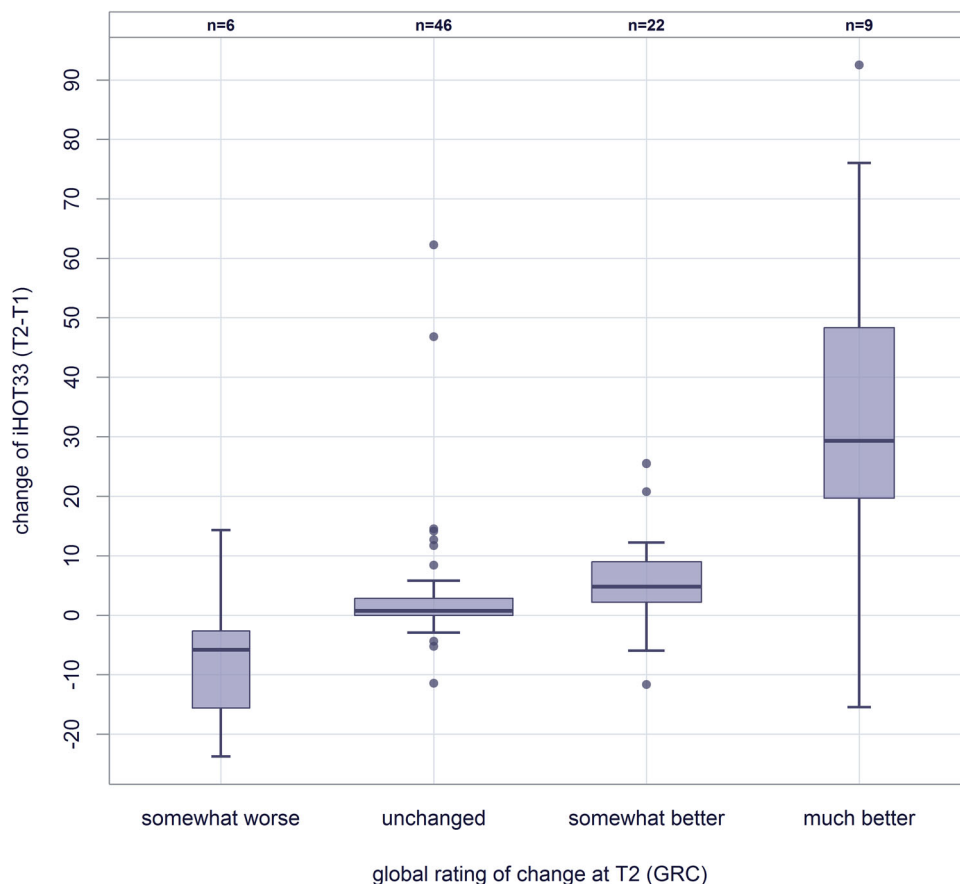


Table 4 Subscale analysis for G-iHOT33

Subscale	Cronbach's alpha for iHOT33 subscales	ICC for iHOT33 subscales	Floor effect (%)	Ceiling effect (%)	Score subscale correlation	Global functional rating correlation
Symptoms/function	0.96 (0.93, 0.99)	0.79 (0.66, 0.88)	0.0	9.6	0.82 (HOS)	0.62
Sports activity	0.93 (0.85, 1)	0.94 (0.89, 0.97)	2.4	7.2	0.75 (HOS)	0.69
Occupational	0.88 (0.76, 1.00)	0.88 (0.80, 0.93)	4.8	10.8	–	0.62
Social interactions	0.91 (0.84, 0.99)	0.88 (0.80, 0.93)	1.2	9.6	0.66 (EQ5D)	0.66
Total ($n = 83$)	0.97 (0.96, 0.99)	0.88 (0.80, 0.99)	0.0	6.0	0.86 (HOS)	0.72

becoming more and more important to reflect the patient's view of the postoperative outcome and limitations in everyday life. For patients with hip fractures, osteoarthritis or those undergoing hip replacement several established PRO already exist. Recently, some questionnaires were developed to evaluate the postoperative outcome in this cohort of young, physically active patients. These questionnaires mainly focus on symptoms and function in everyday life [2–4]. Most of them, however, do not give a comprehensive picture of the patients' views. Some authors have already pointed out a discrepancy between functional results and patient satisfaction in patients undergoing hip arthroscopy [5–7, 30, 31]. Social, emotional, and

occupational factors might also play an important role in the patients' assessment of the therapy. Therefore, the Multicenter Arthroscopy of the Hip Outcomes Research Network (MAHORN) has recently developed an outcome measurement instrument including questions on social, emotional, and occupational limitations. The iHOT has also been cross-culturally adapted into Spanish, Portuguese, and Swedish [32–34]. Recent comparative studies have shown good results for most of the psychometric properties of the iHOT33 [4, 9, 12, 18, 33]. Our study also showed a high level of reliability and validity for the German version of the iHOT-33. However, some authors have made critical remarks about the methodical quality of

the development study of the iHOT33 and advised the separate validation of the iHOT33 subscales [4, 13, 35, 36].

Study design and population

Our demographic data are comparable to other studies on young active patients generally undergoing hip arthroscopy with an average age around 35 years [6–8, 12, 29, 30]. To get a more heterogeneous patient sample, we did not pre-select patients according to their diagnosis or intended treatment. Aiming for validation data for various hip diseases, we included all patients with a hip disorder and an activity level greater than 4 on the Tegner activity scale. Compared to the original publication of the iHOT33 by Mohtadi [9], we chose more liberal inclusion criteria sparing a limitation of age. We nonetheless excluded patients with a disorder of the back or the contralateral lower extremity or a mental disorder to avoid confounding. The number of patients included in our study is according to previous recommendations [23].

The translation process was conducted according to the guidelines of the American Academy of Orthopedic Surgeons (AAOS) Outcomes Committee [14]. The period of time between test and retest was chosen to be a minimum of 2 weeks as recommended in the COSMIN checklist [22]. The validation was carried out following the complete COSMIN checklist [10, 22]. Along with the prospective multicenter design, the study meets high methodological standards with a level of evidence Ib.

Reliability

The excellent correlation coefficient for Cronbach's alpha outlines the quality of the German iHOT33 and confirms the results of prior validation studies on the iHOT33 [9, 12, 18, 33]. Accordingly, an ICC of 0.88 confirmed good test–retest reliability. Low values for measurement error and smallest detectable change (SDC) indicate that small clinical changes can be detected not only at group level but also at the individual level [23]. Harris-Hayes et al. [13] contended that one current limitation of the iHOT33 is that the subscales concerning symptoms/function, sports activity, social, and occupational issues were not validated separately. The present study is the first to give a Cronbach's alpha for each subscale and therefore confirm a positive rating for internal consistency [9, 12, 18, 33].

Validity

For the evaluation of the construct validity, the HOS and EQ-5D seemed most appropriate because they are applicable in this patient population and validated in the

German language [17, 20, 21]. The HOS was also developed for patients undergoing hip arthroscopy [17]. Therefore, it seemed a viable instrument to evaluate construct validity. We determined a sufficient level of activity by including only patients with an activity level greater than four on the modified Tegner Activity Scale (mTAS) as described by Mohtadi et al. [9]. Because the mTAS is solely derived from the level of sports activity, it is rather robust to smaller changes of the medical condition. Therefore, we expected rather low correlations between the functional hip scores and the mTAS.

Responsiveness and Interpretability

In our population, the minimal important change (MIC) according to the method of Norman et al. [25] was 12.7. This indicates a reasonable ability to transform a qualitative effect into a quantitative score while providing a high discriminatory power. Using this method is controversial in the literature [4, 7, 19, 37, 38]. Norman et al. [25] suggested an estimation of the MIC by division of the standard deviation by two. This method was derived from the effect size when undergoing clinical intervention. Based on Terwee et al. [23], a positive rating for responsiveness can be assumed when SDC is greater than MIC. With an anchor-based method, the patient's report a current state of health targeting for their personal expectations [4, 37, 38]. Therefore, calculation of MIC can be problematic in absence of a therapeutic gold standard. The COSMIN checklist does not contain any recommendations on the estimation of the MIC. Still, there is no consensus about how to assess the MIC. Although the 95 % CI of the “unchanged” and “somewhat better” group is wide, a remarkable difference exists only within a small range of 0.21–0.63. The cutoff point that separates the “unchanged” group from the “somewhat better” group has to be outside of the 95 % CI of the “unchanged” group. It should also represent the smallest acceptable effect size for the “somewhat better” group. Accordingly, an estimated ES of around 0.6 seems reasonable to approximate the true ES of the iHOT33. Consequently, the minimal important change that reflects a clinically relevant improvement is approximately ten points on the iHOT33. Defining the MIC of the iHOT-33 is a main aspect of this study.

Our findings according to the global rating of change also confirm a strong correlation between patient perception and score result of the G-iHOT33. Nevertheless, a change of “somewhat better” or “somewhat worse” was not statistically significant compared to a steady state.

None of the iHOT33 subscales showed floor- or ceiling-effects. These results are concordant with previous results on interpretability of the iHOT33 [4, 9, 12, 18, 33].

Limitations

Despite good results concerning validity, reliability, and responsiveness of the G-iHOT33, there are some limitations to this study. Firstly, there was no additional score to correlate with the occupational subscale of the iHOT33. Therefore, the occupational subscale could only be correlated with the patient's subjective rating showing reasonable correlation.

Another limitation is that due to a lack of a therapeutic gold standard, the minimal important change could only be estimated. To date, the only study providing longitudinal results on the iHOT is Mas Martinez et al. [8]. They report on short-term outcomes after hip arthroscopy in FAI with a minimum follow-up of 12 months. Unfortunately, data concerning minimal clinically important change are not specified in this study. Further prospective studies on longitudinal measurement properties of the iHOT33 are needed.

Conclusion

The German version of the iHOT33 provides good validity, reliability, and responsiveness for the functional evaluation of physically active patients with a hip disorder. This is the first study to present validation results for the iHOT33 subscales. Another main aspect of this study is the definition of the minimal important change (MIC) which is 10 points on the iHOT-33 scale. The COSMIN checklist is a feasible guideline to assess psychometric properties of patient-reported outcome measurements.

Acknowledgments The authors would like to thank Mr. Daniel Popp for his assistance in data collection.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Klassbo M, Larsson E, Mannevik E (2003) Hip disability and osteoarthritis outcome score. An extension of the Western Ontario and McMaster Universities Osteoarthritis Index. *Scand J Rheumatol* 32(1):46–51 (PubMed PMID: 12635946)
- Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW (1988) Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *J Rheumatol* 15(12):1833–1840
- 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(16):1186–1196. doi:10.1136/bjsm.2009.060889
- Thorborg K, Tjissen M, Habets B, Bartels EM, Roos EM, Kemp J et al (2015) Patient-Reported Outcome (PRO) questionnaires for young to middle-aged adults with hip and groin disability: a systematic review of the clinimetric evidence. *Br J Sports Med*. doi:10.1136/bjsports-2014-094224
- Leunig M, Ganz R (2014) The evolution and concepts of joint-preserving surgery of the hip. *Bone Joint J* 96-B(1):5–18. doi:10.1302/0301-620X.96B1.32823
- Bogunovic L, Gottlieb M, Pashos G, Baca G, Clohisy JC (2013) Why do hip arthroscopy procedures fail? *Clin Orthop Relat Res* 471(8):2523–2529. doi:10.1007/s11999-013-3015-6
- Mannion AF, Impellizzeri FM, Naal FD, Leunig M (2013) Fulfilment of patient-rated expectations predicts the outcome of surgery for femoroacetabular impingement. *Osteoarthritis Cartilage* 21(1):44–50. doi:10.1016/j.joca.2012.09.013
- Mas Martinez J, Morales-Santias M, Bustamante Suarez Suarezde Puga D, Sanz-Reig J (2014) Hip arthroscopy in males younger than 40 with femoroacetabular impingement: short-term outcomes. *Rev Esp Cir Ortop Traumatol* 58(6):343–350. doi:10.1016/j.recot.2014.05.003
- Mohtadi NG, Griffin DR, Pedersen ME, Chan D, Safran MR, Parsons N et al (2012) 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* 28(5):595–605. doi:10.1016/j.arthro.2012.03.013 (quiz 6–10 e1)
- Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL et al (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(4):539–549. doi:10.1007/s11136-010-9606-8
- Hinman RS, Dobson F, Takla A, O'Donnell J, Bennell KL (2014) Which is the most useful patient-reported outcome in femoroacetabular impingement? Test–retest reliability of six questionnaires. *Br J Sports Med* 48(6):458–463. doi:10.1136/bjsports-2012-092072
- Kemp JL, Collins NJ, Roos EM, Crossley KM (2013) Psychometric properties of patient-reported outcome measures for hip arthroscopic surgery. *Am J Sports Med* 41(9):2065–2073. doi:10.1177/0363546513494173
- Harris-Hayes M, McDonough CM, Leunig M, Lee CB, Callaghan JJ, Roos EM (2013) Clinical outcomes assessment in clinical trials to assess treatment of femoroacetabular impingement: use of patient-reported outcome measures. *J Am Acad Orthop Surg* 21(Suppl 1):S39–S46. doi:10.5435/JAAOS-21-07-S39
- Beaton DE, Bombardier C, Guillemin F, Ferraz MB (2000) Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine* 25(24):3186–3191
- Tegner Y, Lysholm J (1985) Rating systems in the evaluation of knee ligament injuries. *Clin Orthop Relat Res* 198:43–49
- Martin RL, Kelly BT, Philippon MJ (2006) Evidence of validity for the hip outcome score. *Arthroscopy* 22(12):1304–1311. doi:10.1016/j.arthro.2006.07.027
- Naal FD, Impellizzeri FM, Miozzari HH, Mannion AF, Leunig M (2011) The German Hip Outcome Score: validation in patients undergoing surgical treatment for femoroacetabular impingement. *Arthroscopy* 27(3):339–345. doi:10.1016/j.arthro.2010.07.021
- Griffin DR, Parsons N, Mohtadi NG, Safran MR, Multicenter Arthroscopy of the Hip Outcomes Research N (2012) A short version of the International Hip Outcome Tool (iHOT-12) for use

- in routine clinical practice. *Arthroscopy* 28(5):611–616. doi:10.1016/j.arthro.2012.02.027 (quiz 6–8)
19. Ramisetty N, Kwon Y, Mohtadi NG (2015) Patient-reported outcome measures for hip preservation surgery—a systematic review of the literature. *J Hip Preserv Surg*. doi:10.1093/jhps/hnv002 (Epub 06.02.2015)
 20. Rabin R, de Charro F (2001) EQ-5D: a measure of health status from the EuroQol Group. *Ann Med* 33(5):337–343
 21. Greiner W, Claes C, Busschbach JJ, von der Schulenburg JM (2005) Validating the EQ-5D with time trade off for the German population. *Eur J Health Econ* 6(2):124–130
 22. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL et al (2012) COSMIN checklist manual. <http://www.cosmin.nl/images/upload/files/COSMIN%20checklist%20manual%20v9.pdf>
 23. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J et al (2007) Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 60(1):34–42. doi:10.1016/j.jclinepi.2006.03.012
 24. 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(6):478–491. doi:10.1136/bjsm.2010.080937
 25. Norman GR, Sloan JA, Wywich KW (2003) Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Med Care* 41(5):582–592. doi:10.1097/01.MLR.0000062554.74615.4C
 26. Tibor LM, Leunig M (2012) Labral resection or preservation during FAI treatment? A systematic review. *HSS J* 8(3):225–229. doi:10.1007/s11420-012-9294-8
 27. Tijssen M, van Cingel R, Willemsen L, de Visser E (2012) Diagnostics of femoroacetabular impingement and labral pathology of the hip: a systematic review of the accuracy and validity of physical tests. *Arthroscopy* 28(6):860–871. doi:10.1016/j.arthro.2011.12.004
 28. Harris JD, McCormick FM, Abrams GD, Gupta AK, Ellis TJ, Bach BR Jr et al (2013) Complications and reoperations during and after hip arthroscopy: a systematic review of 92 studies and more than 6000 patients. *Arthroscopy* 29(3):589–595. doi:10.1016/j.arthro.2012.11.003
 29. Lynch TS, Terry MA, Bedi A, Kelly BT (2013) Hip arthroscopic surgery: patient evaluation, current indications, and outcomes. *Am J Sports Med* 41(5):1174–1189. doi:10.1177/0363546513476281
 30. McCarthy JC, Jarrett BT, Ojeifo O, Lee JA, Bragdon CR (2011) What factors influence long-term survivorship after hip arthroscopy? *Clin Orthop Relat Res* 469(2):362–371. doi:10.1007/s11999-010-1559-2
 31. 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(3):544–550. doi:10.1177/0363546510387263
 32. Polesello GC, Godoy GF, Trindade CA, de Queiroz MC, Honda E, Ono NK (2012) Translation and cross-cultural adaptation of the International Hip Outcome Tool (iHOT) into Portuguese. *Acta Ortop Bras* 20(2):88B–92B. doi:10.1590/S1413-78522012000200006
 33. Jonasson P, Baranto A, Karlsson J, Sward L, Sansone M, Thomee C et al (2014) A standardised outcome measure of pain, symptoms and physical function in patients with hip and groin disability due to femoro-acetabular impingement: cross-cultural adaptation and validation of the international Hip Outcome Tool (iHOT12) in Swedish. *Knee Surg Sports Traumatol Arthrosc* 22(4):826–834. doi:10.1007/s00167-013-2710-x
 34. Ruiz-Iban MA, Seijas R, Sallent A, Ares O, Marin-Pena O, Muriel A et al (2015) The international Hip Outcome Tool-33 (iHOT-33): multicenter validation and translation to Spanish. *Health Qual Life Outcomes* 13:62. doi:10.1186/s12955-015-0255-z
 35. Thorborg K, Roos EM, Christensen R, Petersen J, Holmich P (2012) The iHOT-33: how valid is it? *Arthroscopy* 28(9):1194–1195. doi:10.1016/j.arthro.2012.07.003 (author reply 5–6)
 36. Thorborg K, Holmich P (2014) Measurement qualities of hip and groin outcome scores: letter to the editor. *Am J Sports Med* 42(1):NP7–NP10. doi:10.1177/0363546513514718
 37. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC (2012) Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res* 21(4):651–657. doi:10.1007/s11136-011-9960-1
 38. Tubach F, Ravaud P, Martin-Mola E, Awada H, Bellamy N, Bombardier C et al (2012) Minimum clinically important improvement and patient acceptable symptom state in pain and function in rheumatoid arthritis, ankylosing spondylitis, chronic back pain, hand osteoarthritis, and hip and knee osteoarthritis: results from a prospective multinational study. *Arthritis Care Res* 64(11):1699–1707. doi:10.1002/acr.21747