

RESEARCH

Open Access



Validation of the Brazilian version of the Hip Outcome Score (HOS) questionnaire

Rafaela Maria de Paula Costa^{1,5*}, Themis Moura Cardinot², Letícia Nunes Carreras Del Castillo Mathias^{1,5}, Gustavo Leporace³ and Liszt Palmeira de Oliveira^{4,5}

Abstract

Background: The Hip Outcome Score (HOS) was developed to evaluate physically active patients with hip disease but without severe degenerative change. A translation and cultural adaptation into Brazilian Portuguese was previously conducted. The aim of this study was to validate the Brazilian version of the HOS (HOS-Brazil) among a group of physically active patients with a diagnosis of femoroacetabular impingement (FAI) or greater trochanteric pain syndrome (GTPS).

Methods: The following questionnaires were applied: the HOS-Brazil; the validated Brazilian versions of the Nonarthritic Hip Score (NAHS) and the 12-Item Short-Form Health Survey (SF-12). The psychometric properties analyzed with regard to the validation process were reliability and validity. Internal consistency and intra-rater test-retest reliabilities were analyzed using Cronbach's alpha and the intraclass correlation coefficient (ICC) statistical tests based on test-retest agreement. Construct and content validities were examined using Pearson's correlation coefficient. Content validity was also analyzed based on evidence of floor, ceiling, or both types of effects from the questionnaires.

Results: A total of 70 male and female patients were selected, aged between 19 and 70 years old. The internal consistency and intra-rater test-retest reliability values were high (Cronbach's $\alpha > 0.9$; ICC > 0.9). The questionnaire showed acceptable convergent ($r > 0.7$) and divergent ($r < 0.4$) validities. No floor or ceiling effects were observed.

Conclusion: The HOS-Brazil was validated. Additional studies are underway to evaluate its responsiveness.

Keywords: Questionnaires, Hip, Hip outcome score, Validity, Reliability

Background

The Hip Outcome Score (HOS) is an instrument used to assess patients with hip disorders who are young, physically active, or both but who do not have severe degenerative abnormalities. Other hip assessment instruments do not address this population with the same degree of specificity [1–5].

Martin developed the HOS in the United States of America (USA) in 2005 to assess patients with acetabular labral tears who were physically active, young, or both [1]. The instrument was validated using two

groups: individuals receiving hip arthroscopy and those with acetabular labral tears [6, 7].

Most quality of life instruments and orthopedic assessments were originally developed in English [1, 2, 8, 9]. For these instruments to be used across cultures and languages, several steps of translation and cross-cultural adaptation should be accomplished. These steps should be followed by validation to determine whether the new instrument conserves the psychometric characteristics of the original [10–12].

The standardized set of instructions for the translation and cultural adaptation of quality of life assessments includes five steps: translation, back-translation, review by a committee pretesting and final translation. Guillemin et al. [10] first described these criteria, which were later revised by Beaton et al. [11]. Following translation and cultural

* Correspondence: rafaelacosta87@gmail.com

¹Medical Sciences, State University of Rio de Janeiro, Rio de Janeiro, Brazil
⁵Orthopedics Service, Pedro Ernesto University Hospital, State University of Rio de Janeiro, Rio de Janeiro, RJ, Brazil

Full list of author information is available at the end of the article



adaptation, the measurement (i.e., psychometric) properties of instruments should be tested (i.e., validated) [12–14].

The psychometric properties usually analyzed for the purpose of validation are reliability, validity, and responsiveness [12, 13]. These properties were standardized by researchers who developed consensus-based guidelines for the selection of measurement properties for the validation of health instruments (i.e., the *Consensus-based Standards for the Selection of Health Measurement Instruments*; COSMIN) to assess the methodological quality of studies that use these measurement properties [13–15].

The present study sought to validate the Brazilian version of HOS (HOS-Brazil) using a group of physically active patients diagnosed with femoroacetabular impingement (FAI) or greater trochanteric pain syndrome (GTPS). This validated questionnaire will provide doctors and other health-care providers in Brazil with a more specific instrument to assess this population of patients. Importantly, the hip research group at Pedro Ernesto University Hospital, State University of Rio de Janeiro (Hospital Universitário Pedro Ernesto, Universidade do Estado do Rio de Janeiro; HUPE/UERJ) previously translated and culturally adapted the HOS [16].

Methods

The HUPE/UERJ research ethics committee approved the present study (CEP/HUPE no. 2674). The participants were informed of the study aims and methods before signing an informed consent document.

Patient selection

A total of 70 male and female patients who were literate and physically active who reported hip pain and were diagnosed with either FAI or GTPS (as confirmed by radiograph, tomography, or magnetic resonance imaging) were selected. The participants were recruited from the Orthopedic Institute of Tijuca, a private hip outpatient clinic in Rio de Janeiro. Data were collected between December 2015 and June 2016.

Patients were excluded if they showed visual or cognitive disorders that impaired the reading and interpretation of the questions; hip arthrosis, characterized by a minimum joint space of < 1.5 mm and a severe limitation in hip range of motion [17]; or incomplete responses to the questionnaires on day 1 and 48 h after the first application.

Study protocol

The study protocol included completing the identification form, which was composed of the clinical characteristics of the patients and the application of three quality of life assessments: the HOS-Brazil as well as the Brazilian-validated versions of 12-Item Short-

Form Health Survey (SF-12) and the Nonarthritic Hip Score (NAHS) [16, 18, 19]. The participants were instructed to complete all three questionnaires (1st application or test). Approximately 48 h later, they completed only the HOS-Brazil via e-mail (2nd application or retest).

HOS-Brazil

The HOS is a self-report questionnaire composed of 28 questions divided into two subscales: Activities of Daily Living (ADLs; 19 items) and Sports (nine items) [1, 16]. The total score of each subscale ranges from 0 to 100, where higher scores denote better hip function. The scores for each subscale were calculated separately [6].

The response options are the same for all 28 items and given specific scores that are added to the end of the assessment. Responses to the 19 items of the ADL subscale are scored from 0 to 4, where 4 is “No difficulty at all” and 0 is “Unable to perform.” The scores of the individual items are added to obtain the total score, which is then multiplied by 4 to generate the highest potential score. Assuming the patients respond to all 19 items, the highest possible score is 76. The total score obtained is divided by 76. The resulting value is multiplied by 100 to express the score as a percentage. The nine items of the Sports scale are calculated in the same way, and the highest possible score is 36. A higher final score represents a better level of physical functioning with regard to both the ADL and Sports subscales [6].

In addition, the HOS includes two questions regarding how respondents rate their current level of functioning during ADLs and sports from 0 to 100 as well as one qualitative question asking them to rate their current level of functioning (normal, nearly normal, abnormal, or severely abnormal). The responses to these three questions are not considered in the HOS final score [20].

Psychometric properties

To validate the psychometric properties of the HOS-Brazil, its reliability and validity were assessed according to the COSMIN checklist [13–15].

Reliability is a psychometric property that measures the degree to which a questionnaire is free from measurement error; furthermore, this process establishes whether the scores remain similar after repeated application to the same sample on a different occasion and without the influence of treatment. The reliability of the HOS-Brazil was assessed based on the following properties: internal consistency, intrarater test-retest reliability, measurement error, and concordance [12, 14, 21].

Internal consistency assesses the ability of a set of questions to measure a similar concept. Test-retest

reliability is a measurement property that assesses the ability of a questionnaire to yield similar results when the same respondents are assessed on a different occasion without undergoing any change in health. Concordance is related to systematic and random errors in the respondents' scores that are not attributed to true changes in the construct to be measured [12, 14].

The reliability of the HOS-Brazil was investigated using a sample of 70 patients who responded to the questionnaire twice with a 48-h interval. In between applications, no new medications, therapies, or procedures were introduced that were likely to induce rapid changes to the patients' clinical conditions. The interval between the test and retest was selected based on two criteria: The period of time was held long enough for the respondents not to remember their previous responses but was simultaneously brief enough for no changes to occur to the patients' clinical conditions [12, 14].

The psychometric property of validity concerns the degree of instrument precision (i.e., whether it conserves the precision of the concept that it intends to measure). Validity assesses whether a new instrument retains the characteristics of the original version and is composed of three measurement properties: construct validity, content validity, and criterion validity [12, 14].

Construct validity corresponds to the degree to which an instrument's scores are consistent with the hypotheses based on the assumption that the instrument measures the intended construct. Content validity estimates the degree to which the content of a measurement instrument is considered as an adequate reflection of the construct to be measured. Criterion validity determines the degree to which an instrument's score adequately reflects the instrument considered as a "gold standard." Criterion validity was not assessed in the present study because, according to the COSMIN formulators, no health assessment instrument is considered as a "gold standard" [12, 14, 21]. Therefore, the validity of the HOS-Brazil was assessed based on only construct and content validity.

Statistical analyses

The intraclass correlation coefficient (ICC) and Pearson's correlation coefficient were used to analyze test-retest reliability [22, 23]. Internal consistency was assessed using Cronbach's alpha [24, 25]. This statistical technique is based on the number of items and their homogeneity within a scale. A paired-sample Student's *t*-test was used to compare the scores obtained for the first and second applications of the HOS-Brazil [23].

Measurement error was calculated based on the standard error of measurement (SEM) and minimal clinically important difference (MCID). SEM was calculated by multiplying the square root of 1 minus

ICC times the standard deviation of the scores obtained for the first application of the HOS-Brazil. The MCID was calculated by multiplying the SEM times 1.96, which is equivalent to the *z*-score of the 95% confidence interval and the square root of 2 [12, 26].

Concordance was assessed based on the graphical representation of the measurement error between test and retest via Bland-Altman and concordance-survival plots [27–29]. The former quantifies concordance using limits of concordance based on the means of the test and retest as well as the difference between both assessments. These statistical limits are calculated using the mean and standard deviation of the differences. A linear regression curve of the Bland-Altman plot was modeled to assess the presence of proportional bias [27, 28]. The independent variable (*x*-axis) used for the linear regression was the mean of both assessments, and the dependent variable (*y*-axis) was the difference between both assessments. The null hypothesis stated that the slope of the regression line would not differ from zero. Proportional bias alludes to a situation in which the difference between the two measurements is not constant across the full range of possible scores as indicated by the *p*-value obtained for regression analysis (*p* < 0.05). If the difference between the scores obtained at two measurements is constant and independent of the scores' magnitude, then it is described as a fixed bias [27, 28].

Construct validity, both convergent and divergent, was assessed using Pearson's correlation coefficient. The HOS-Brazil was compared with the NAHS and SF-12, which have already been validated for Brazilian Portuguese. The aim of the construct validity assessment was to investigate the convergence and divergence of the HOS-Brazil relative to the NAHS and SF-12 [23]. Content validity was assessed based on the presence of completed questionnaires scored as zero or 100 (maximum score), i.e., floor or ceiling effects, respectively [30].

A descriptive statistical analysis was performed to characterize the study sample. The psychometric properties of reliability and validity were analyzed using GraphPad Prism software, version 7.00 for Windows (GraphPad Software, La Jolla, California, USA). The significance level was set at 0.05.

Results

Patient characteristics

Of the 70 selected patients, 46 (65.7%) were female, and 24 (34.3%) were male. The average age was 42.9 ± 12.9 years old (range: 19 to 70 years old). A total of 44 (62.9%) patients were diagnosed with GTPS, and 26 (37.1%) were diagnosed with FAI.

Questionnaire results

The scores on the three applied questionnaires ranged from 2 to 99; higher scores denoted better quality of life (SF-12) and hip function (NAHS and HOS-Brazil). Table 1 describes the means, standard deviations, minimums, and maximums associated with the applied questionnaires.

Psychometric properties

A. Reliability

1. Internal consistency

For the first application of the HOS-Brazil, Cronbach's alphas were 0.95 and 0.92 for the ADL and Sports subscales, respectively (Table 2). The elimination of any isolated question did not significantly change the Cronbach's alpha value for any subscale; therefore, no questions were eliminated from the HOS-Brazil.

2. Intra-rater test-retest reliability

The value obtained was 0.99 for both subscales, with 95% confidence intervals (95% CIs) of 0.986–0.995 and 0.990–0.996 for the ADL and Sports subscales, respectively (Table 2).

3. Measurement error and concordance

Paired-samples Student's t-tests did not show significant differences in the average test-retest scores for either the ADL ($p = 0.84$) or Sports ($p = 0.82$) subscales. The correlation between the test and retest scores was 0.992 (95% CIs = 0.986–0.996, $P < 0.0001$) for the ADL subscale and 0.994 (95% CIs = 0.990–0.996, $P < 0.0001$) for the Sports subscale.

Concordance limits and CIs were analyzed. The Bland-Altman plot showed a mean error of the difference between the test and retest scores of -0.1 for both

Table 2 Psychometric property: Reliability

Internal consistency according to Cronbach's alpha				
Questionnaire	Subscale	Cronbach's alpha		
HOS-Brazil 1st application	ADL	0.95		
	Sports	0.92		
Intra-rater test-retest reliability				
HOS-Brazil test-retest	Subscale	ICC	Lower 95% CI	Upper 95% CI
	ADL	0.992	0.986	0.995
	Sports	0.994	0.990	0.996

HOS-Brazil Brazilian version of Hip Outcome Score, *ADL* activity of daily living, *ICC* intraclass correlation coefficient, *CI* confidence interval

subscales (95% concordance limits = -4.5 to 4.5 for the ADL subscale and -5.3 to 5.2 for the Sports subscale). The two dotted lines represent the upper and lower concordance limits. The P -value of the regression analysis showed that the curve slope did not deviate from zero ($P = 0.26$ for the ADL subscale, $P = 0.14$ for the Sports subscale; Fig. 1a and b).

The concordance and survival plots revealed two findings: a difference of 7 in the ADL subscale scores (Fig. 1c) and of 6 in the Sports subscale scores (Fig. 1d), representing 95% agreement between the test and retest scores.

The SEMs were 1.7 and 1.9 for the ADL and Sports subscale scores, respectively. The calculated MCID was 4.6 for the ADL subscale score and 5.5 for the Sports subscale score.

B. Validity

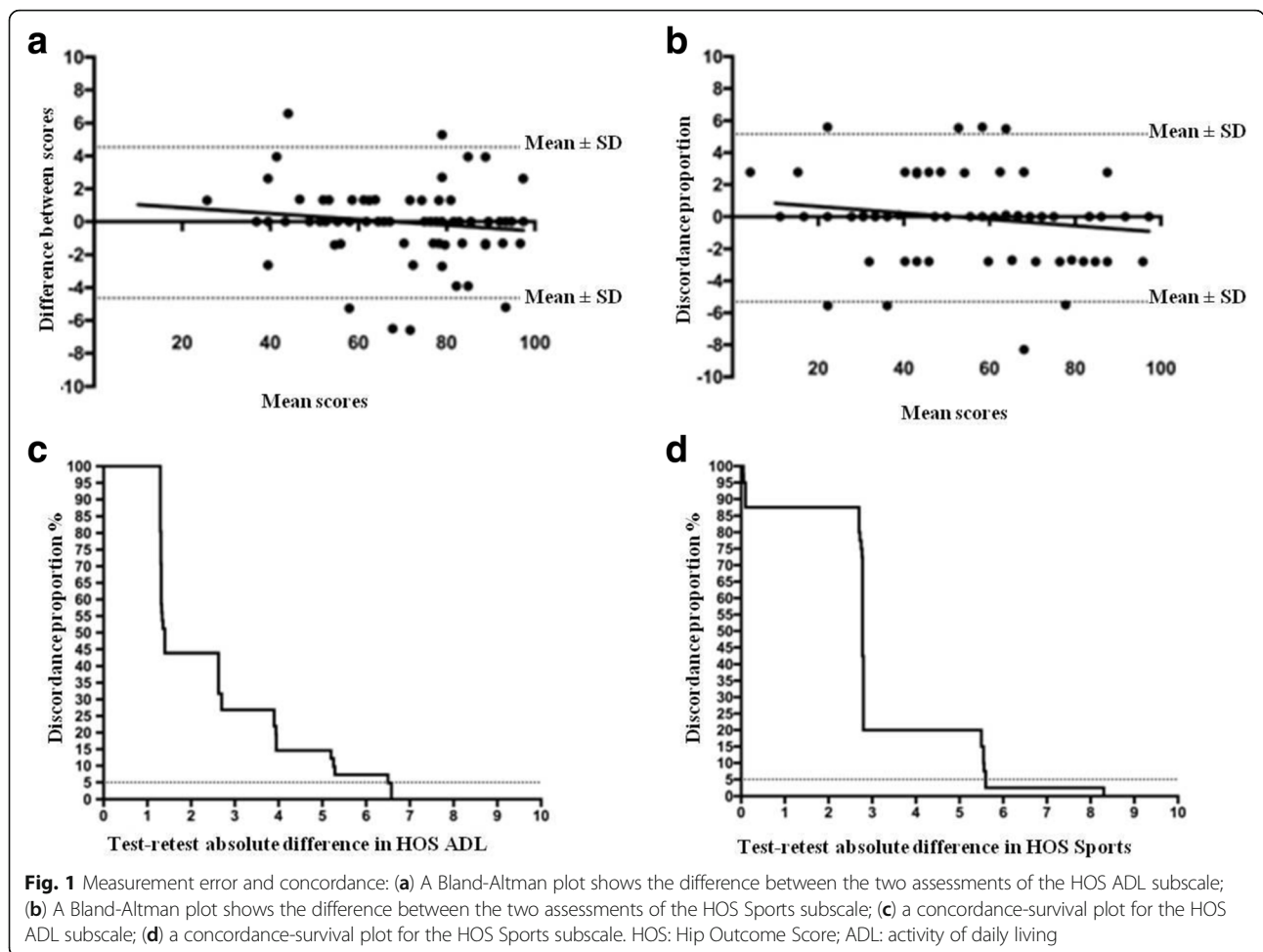
1. Construct validity

Convergent construct validity was estimated based on the correlation between the HOS-Brazil ADL and Sports subscale scores (1st application) and the NAHS total score and SF-12 Physical subscale score. The values of all Pearson's correlation coefficients were over 0.7,

Table 1 The questionnaire scores of 70 patients

Questionnaire	Mean	SD	Minimum score	Maximum score
NAHS 1st application – Total score	62.0	21.8	12.5	96.2
SF-12 1st application, Mental subscale	52.2	9.5	15.8	65.1
SF-12 1st application, Physical subscale	42.9	12.5	20.8	60.9
HOS-Brazil 1st application, ADL subscale	71.1	18.1	25.0	97.4
HOS-Brazil 1st application, Sports subscale	55.1	23.7	2.8	97.2
HOS-Brazil 2nd application, ADL subscale	71.0	17.8	26.3	98.6
HOS-Brazil 2nd application, Sports subscale	55.0	23.3	5.5	97.2

SD standard deviation, *NAHS* Nonarthritic Hip Score, *SF-12* 12-Item Short-Form Health Survey, *HOS-Brazil* Brazilian version of Hip Outcome Score, *ADL* activity of daily living



except for the correlation between the HOS-Brazil Sports subscale and the SF-12 Physical subscale, which was 0.685 (Table 3).

Next, the divergent construct validity between the HOS-Brazil and SF-12 was analyzed. Pearson’s correlation coefficient was calculated to investigate the presence of a correlation between the HOS-Brazil ADL and

Sports subscale scores and the SF-12 Mental subscale score. The results obtained were less than 0.4, indicating a low correlation and, consequently, a lack of convergence (Table 3).

2. Content validity

The obtained content validity was satisfactory; no questionnaire exhibited scores of zero or 100 (maximum score); i.e., floor or ceiling effects were not found with regard to the HOS-Brazil (Fig. 2).

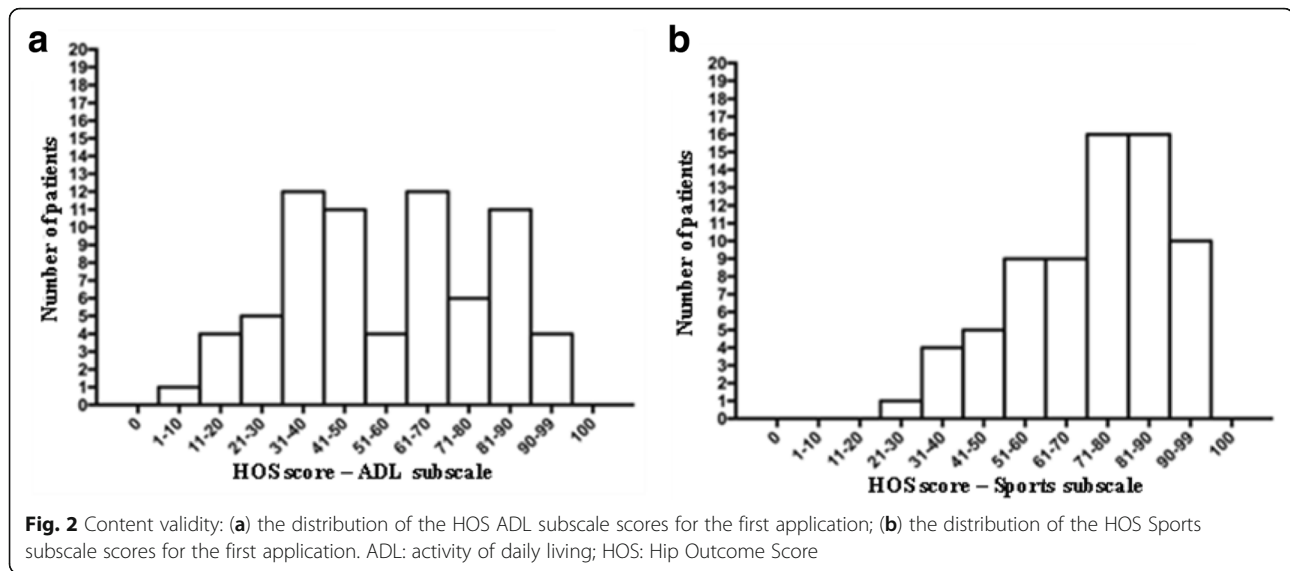
Discussion

The HOS is a quality of life assessment instrument specific to people with hip disease without severe degenerative abnormalities that was originally developed in English [1]. The HOS was translated and cross-culturally adapted into German, Korean, Spanish, and Brazilian Portuguese [16, 20, 31, 32]. The German, Korean, and Spanish versions are already validated for use in their corresponding countries [20, 31, 32]. A group of physically active individuals diagnosed with FAI or GTPS was

Table 3 Psychometric property: Validity

Questionnaires (subscales)	Pearson’s r
Pearson’s correlation coefficient for convergent validity	
HOS-Brazil (ADL subscale) vs. NAHS (total)	0.874***
HOS-Brazil (ADL subscale) vs. SF-12 (Physical subscale)	0.744***
HOS-Brazil (Sports subscale) vs. NAHS (total)	0.789***
HOS-Brazil (Sports subscale) vs. SF-12 (Physical subscale)	0.685***
Pearson’s correlation coefficient for divergent validity	
HOS-Brazil (ADL subscale) vs. SF-12 (Mental subscale)	0.346**
HOS-Brazil (Sports subscale) vs. SF-12 (Mental subscale)	0.344**

HOS-Brazil Brazilian version of Hip Outcome Score, ADL activity of daily living, NAHS Nonarthritic Hip Score, SF-12 12-Item Short-Form Health Survey, r Pearson’s correlation coefficient *** $p < 0.0001$; ** $p = 0.003$



selected for the present validation study of the HOS-Brazil.

A total of 70 patients with an average age of 42.9 years were assessed for the present validation of the HOS-Brazil. In the validation studies of the German, Korean, and Spanish versions, the numbers of patients varied from 60 to 100, and their average age varied from 33 and 45 years [20, 31, 32]. Therefore, the average age of the patients included in the HOS-Brazil validation study was similar to that of the German, Korean, and Spanish studies. Females were more prevalent in the population included for the HOS-Brazil study, which matches the sample recruited to validate the original HOS as well as assess its reliability and responsiveness [6, 33].

The values for internal consistency were high (>0.9) for both scales (0.95 and 0.92 for ADL and Sports, respectively). According to Hair et al. [25] the minimum recommended value is 0.7; 0.8 to 0.9 is rated as moderate to high; and >0.9 is high. Therefore, the questions included in both subscales likely provide a clear reflection of the subject they investigate, which is indicative of sufficient homogeneity among all the items [24, 25]. These internal consistency findings are similar to those obtained in the validation of the original version (and other versions) of the HOS [6, 20, 31, 32].

The HOS-Brazil exhibited excellent intra-rater test-retest reliability, with an ICC of 0.992 for the ADL subscale and 0.994 for the Sports subscale. An ICC ranging from 0.4 to 0.75 is considered satisfactory, whereas values ≥ 0.75 are considered excellent [32]. The analysis showed that all of the ICC values resulting from the comparison of the first and second application were over 0.9 for both subscales. This finding indicates high intra-rater test-retest reliability and shows that the HOS-Brazil

is reliable. The time interval between the test and retest was 48 h.

In the validation studies of the German, Korean, and Spanish versions of the HOS, the median time between assessments was 10 to 21 days [20, 31, 32]. Perhaps these intervals were not brief enough to ensure that no changes occurred to the clinical conditions of patients. In addition, it is difficult to assert that the patients did not receive any therapeutic support during that time. Nevertheless, the results regarding the HOS-Brazil were similar to those obtained after assessing the reliability of the original HOS and the aforementioned translated versions, the overly long interval between test and retest notwithstanding [20, 31–33].

The SEMs for the HOS-Brazil ADL and Sports subscale scores were 1.7 and 1.9, respectively, and the MCIDs were 4.6 and 5.3 for these scores, respectively [12, 33]. In the reliability study of the original HOS, the MCID values were ± 4 and ± 8 for the ADL and Sports subscales scores, respectively [33]. For the German version, the SEMs were ± 4 and ± 8 for these scores, respectively, whereas the MCIDs were ± 11 and ± 22 , respectively [31]. For the Spanish version, the SEMs were ± 5.1 and ± 8.5 for the ADL and Sports subscales scores, respectively, whereas the MCIDs were 13.7 and 22.8, respectively [20]. The discrepancies in the results of the HOS-Brazil compared with the German, Korean, and Spanish versions might be because of the overly long intervals between the test and retest of the latter. Knowledge of the amount of measurement error contributes to the assessment of the outcomes of surgery or other treatments received by patients as well as indicates whether clinical changes relevant to patients occurred [12, 14].

A linear regression curve was created for the Bland-Altman plot to investigate the presence of proportional deviation. The independent variable (x-axis) used in the linear regression analysis was the mean between the two assessments of the HOS-Brazil, and the dependent variable (y-axis) the difference between both assessments. The null hypothesis stated that the slope of the regression line would not differ from zero. Importantly, proportional bias alludes to a situation in which the difference between two measurements is not constant across the full range of possible scores, which demonstrates that the curve slope does not deviate from zero. When the difference in the scores between two assessments is constant, independent from the score's magnitude, it is described as fixed polarization. When the regression line is parallel to the x-axis, it provides a demonstration of a fixed bias. The differences between the values obtained in the two applications of the HOS-Brazil remained constant. This analysis quantifies concordance by constructing concordance limits [27, 28].

An analysis of the convergent validity between the HOS-Brazil ADL subscale and the NAHS showed a strong correlation (0.874). The same was true of the correlation between the HOS-Brazil Sports subscale and the SF-12 Physical subscale (0.744). In addition, the correlation between the HOS-Brazil Sports subscale and the NAHS was strong (0.789). The correlation between the HOS-Brazil Sports subscale and the SF-12 Physical subscale was only moderate (0.685). These significant (strong and moderate, respectively) correlations between the HOS-Brazil ADL and Sports subscales and the NAHS and the SF-12 Physical scale allows us to infer that the HOS-Brazil subscales are convergent with the scores of the other two instruments. The highest values correspond to the correlation between the HOS-Brazil and the NAHS, which shows that both instruments exhibit similar characteristics. This finding might be explained by the fact that the NAHS is also an instrument specific to hip disease, and its questions investigate pain, mechanical symptoms, function, and activity, whereas the SF-12 is a generic quality of life questionnaire.

An analysis of the divergent validity between the HOS-Brazil ADL and Sports subscales and the SF-12 Mental scale yielded values of 0.346 and 0.344, respectively. These are weak, non-significant correlations; on these grounds, the analyzed subscales do not converge. In this case, one might infer that a divergence exists between the scores on the HOS-Brazil ADL and Sports subscales and the SF-12 Mental subscale. As a result, one might conclude that the HOS-Brazil adequately converged and diverged relative to the target construct.

The HOS-Brazil exhibited satisfactory content validity because no questionnaire exhibited floor or ceiling effects. This result is similar to that of the

validation study of the Korean version [32], but differs from the German and Spanish validation studies, which detected floor or ceiling effects [20, 31].

The validation studies of the Spanish and Korean versions of the HOS assessed the responsiveness of patients receiving surgery 6 months after treatment [20, 30]. The present validation study did not assess responsiveness because we did not apply the questionnaire to patients after a long period of time. However, the absence of this analysis does not hinder the validation of the HOS-Brazil. Additional studies assessing its responsiveness are currently in progress.

This aspect was one limitation of the present study, which was due to the lack of the prospective reassessment of patients to evaluate their sensitivity to changes in their quality of life following treatment. This limitation derived from a lack of treatment adherence. Another limitation arises from the fact that all participants were recruited from a single center that was part of a private health network in Rio de Janeiro. Therefore, the results might not reflect the experience of the entire Brazilian population.

Conclusions

The HOS-Brazil was validated using a group of physically active patients diagnosed with FAI or GTPS. The psychometric properties of reliability and validity demonstrated excellent internal consistency, intra-rater test-retest reliability, content validity and construct validity.

The present validation study of the HOS-Brazil shows that this instrument is a valid and reliable quality of life assessments in Brazilian Portuguese. Thus, it will provide doctors, physiotherapists and other healthcare providers in Brazil with the ability to assess physically active patients with hip disorders but without severe degenerative abnormalities.

Authors' contributions

RMdePC Elaboration of the article and critical review; Data collection; Data analysis and interpretation; Approval of the final version to be published. TMC Elaboration of the article and critical revision; Approval of the final version to be published. LNCDCM Critical review; Approval of the final version to be published. GL Data analysis and interpretation; Critical review; Approval of the final version to be published. LPdeO Design and study design; Data collection; Data analysis and interpretation; Critical review; Approval of the final version to be published.

Ethics approval and consent to participate

The HUPE/UERJ research ethics committee approved the present study (CEP/HUPE no. 2674). The participants were informed of the study aims and methods before signing an informed consent document.

Consent for publication

Consent for publication was obtained for this study.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Medical Sciences, State University of Rio de Janeiro, Rio de Janeiro, Brazil.
²Medical Sciences, Rural Federal University of Rio de Janeiro, Seropédica, RJ, Brazil.
³Biomedical Engineering, Federal University of Rio de Janeiro, Rio de Janeiro, RJ, Brazil.
⁴State University of Rio de Janeiro, Rio de Janeiro, RJ, Brazil.
⁵Orthopedics Service, Pedro Ernesto University Hospital, State University of Rio de Janeiro, Rio de Janeiro, RJ, Brazil.

Received: 5 April 2018 Accepted: 19 April 2018

Published online: 24 May 2018

References

- Martin RL. Hip arthroscopy and outcome assessment. *Oper Tech Orthop.* 2005;15(3):290–6.
- Christensen CP, Althausen PL, Mittleman MA, Lee JA, McCarthy JC. The nonarthritic hip score: reliable and validated. *Clin Orthop Relat Res.* 2003;406:75–83.
- Klässbo M, Larsson E, Mannevik E. Hip disability and osteoarthritis outcome score. An extension of the western Ontario and McMaster universities osteoarthritis index. *Scand J Rheumatol.* 2003;32(1):46–51.
- Mohtadi NG, Griffin DR, Pedersen ME, Chan D, Safran MR. 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(5):595–605.
- Naal FD, Miozzari HH, Kelly BT, Magennis EM, Leunig M, Noetzi HP. The hip sports activity scale (HSAS) for patients with femoroacetabular impingement. *Hip Int.* 2013;23(2):204–11.
- Martin RL, Kelly BT, Philippon MJ. Evidence of validity for the hip outcome score. *Arthroscopy.* 2006;22(12):1304–11.
- Martin RL, Philippon MJ. Evidence of validity for the hip outcome score in hip arthroscopy. *Arthroscopy.* 2007;23(8):822–6.
- Harris WH. Traumatic arthritis of the hip after dislocation and acetabular fractures: treatment by mold arthroplasty. *J Bone Joint Surg Am.* 1969; 51:737–55.
- Marx FC, Oliveira LM, Bellini CG, Ribeiro MCC. Tradução e validação cultural do questionário algofuncional de Lequesne para osteoartrite de joelhos e quadris para a língua portuguesa. *Rev Bras Reumatol.* 2006;46(4):253–60.
- 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(12):1417–32.
- Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine.* 2000; 25(24):3186–91.
- Scholtes VA, Terwee CB, Poolman RW. What makes a measurement instrument valid and reliable? *Injury. Int J Care Injured.* 2011;42:236–40.
- Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, et al. Protocol of the COSMIN study: CONsensus-based standards for the selection of health measurement instruments. *BMC Med Res Methodol.* 2006;6:2.
- Mokkink LB, Prinsen CAC, Bouter LM, De Vet HCW, Terwee CB. The CONsensus-based standards for the selection of health measurement instruments (COSMIN) and how to select an outcome measurement instrument. *Braz J Phys Ther.* 2016;20(2):105–13.
- Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick 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.
- Oliveira LP, Cardinot TM, Del Castillo LNC, Queiroz MC, Polesello GC. Translation and cultural adaptation of the hip outcome score to the Portuguese language. *Rev Bras Ortop.* 2014;49(3):297–304.
- Croft P, Coope RC, Wickham C, Coggon D. Defining osteoarthritis of the hip for epidemiologic studies. *Am J Epidemiol.* 1990;132(3):514–22.
- Silveira MF, Almeida JC, Freire RS, Haikal DS, Martins AEBL. Propriedades psicométricas do instrumento de avaliação da qualidade de vida: 12-item health survey (SF-12). *Ciênc saúde coletiva.* 2013;18(7):1923–31.
- Del Castillo LN, Leporace G, Cardinot TM, Levy RA, Oliveira LP. Translation, cross-cultural adaptation and validation of the Brazilian version of the nonarthritic hip score. *São Paulo Med J.* 2013;131(4):244–51.
- Seijas R, Sallent A, Ruiz-Ibán MA, Ares O, Marín-Peña O, Cuéllar R, et al. Validation of the Spanish version of the hip outcome score: a multicenter study. *Health Qual Life Outcomes.* 2014;12:70.
- Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol 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.
- Bartko JJ. The Intraclass correlation coefficient as a measure of reliability. *Psychol Rep.* 1966;19:3–11.
- Vieira S. *Bioestatística: tópicos avançados.* 3ed ed. Rio de Janeiro: Elsevier; 2010.
- Cronbach LJ. Coefficient alpha and internal structure of tests. *Psychometrika.* 1951;16:297–334.
- Hair Jr JF, Anderson RE, Tatham RL, Black WC. *Análise multivariada de dados.* In: Porto Alegre: Book. 6ed ed; 2009.
- Bartlett JW, Frost C. Reliability, repeatability and reproducibility: analysis of measurement errors in continuous variables. *Ultrasound Obstet Gynecol.* 2008;31:466–75.
- Altman DG, Bland JM. Measurement in medicine: the analysis of method comparison studies. *Underst Stat.* 1983;32:207–17.
- Bland JM, Altman DG. Statistical methods of assessing agreement between two methods of clinical measurement. *Lancet.* 1986;1:307–10.
- Luiz RR, Costa AJL, Kale PL, Werneck GL. Assessment of agreement of a quantitative variable: a new graphical approach. *J Clin Epidemiol.* 2003; 56(10):593–7.
- Everitt BS, Skrondal A. *The Cambridge dictionary of statistics.* 4th ed. New York: Cambridge University Press; 2010.
- Naal FD, Impellizzeri FM, Miozzari HH, Mannion AF, Leunig M. The German hip outcome score: validation in patients undergoing surgical treatment for femoroacetabular impingement. *Arthroscopy.* 2011;27(3):339–45.
- Lee YK, Ha YC, Martin RL, Hwang DS, Koo KH. Transcultural adaptation of the Korean version of the hip outcome score. *Knee Surg Sports Traumatol Arthrosc: Official Journal of the ESSKA.* 2014;23(11):3426–31. <https://doi.org/10.1007/s00167-014-2946-0>.
- Martin RL, Philippon MJ. Evidence of reliability and responsiveness for the hip outcome score. *Arthroscopy.* 2008;24(6):676–82.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://www.biomedcentral.com/submissions)

