Original Research Article

HIP International

HIP International
1–10

DOI: 10.1177/1120700020901682
journals.sagepub.com/home/hpi
SAGE

Modified Harris Hip Score as patient-reported outcome measure in osteoarthritic patients: psychometric properties of the Greek version

Sophia Stasi[1], George Papathanasiou[1], Afroditi Diochnou[1], Basiliki Polikreti[1], Antonios Chalimourdas[1] and George A Macheras[1,2]

Abstract

Introduction: This study explored the psychometric properties of the modified Harris Hip Score-Greek version (mHHS-Gr) as a patient-reported outcome (PRO) measure in osteoarthritic hip patients.

Methods: Internal consistency, test-retest reliability and reproducibility were evaluated in 90 patients aged >55 years. Construct validity was tested against Greek versions of the Lower Extremity Functional Scale (LEFS-Greek) and WOMAC Index (WOMAC-Gr), and the Timed Up and Go (TUG) and 9-stairs-ascend/descend (9S-A/D) tests. Known-groups validity was examined using TUG score (cut-off 13.5 s) as an estimate variable. Responsiveness was examined before and 4 weeks after direct anterior minimal invasive surgery.

Results: *Reliability:* Internal consistency was moderate (Cronbach's $a = 0.614$, $p < 0.001$). Test-retest reliability was excellent (ICC = 0.881, 95% CI, 0.824–0.920). *Reproducibility*: Floor and ceiling effects were both 1.1%; measurement error was 3.54 ($p < 0.05$); minimal important change was lower than minimal detectable change.

Validity: mHHS-Gr correlated strongly with both LEFS-Greek and WOMAC-Gr (Pearson's r 0.801 and −0.783, respectively; $p < 0.001$). The questionnaire's correlations with TUG and 9S-A/D were also significant but moderate (Spearman's ρ: −0.547 and −0.575, respectively; $p < 0.001$). Known-groups validity showed that mHHS-Gr scores were significantly higher in participants with TUG < 13.5 seconds than in those with TUG > 13.5 seconds ($p < 0.001$). In ROC analysis, the cut-off point of 52.5 yielded sensitivity 81% and specificity 71%.

Responsiveness: Standardised response mean and Guyatt's responsiveness statistic were greater than 0.8.

Discussion: mHHS-Gr showed significant moderate to excellent reliability, significant moderate to strong validity properties and excellent responsiveness. Overall, mHHS-Gr could be a reliable and valid PRO measure for assessing patients with osteoarthritis of the hip.

Keywords

Osteoarthritis-hip, patient-reported outcome measures, physical-performance measures, reliability, validity

Date received: 18 March 2019; accepted: 13 October 2019

[1]Laboratory of Neuromuscular and Cardiovascular Study of Motion, Physiotherapy Department, Faculty of Health and Care Sciences, University of West Attica, Egaleo, Attica, Greece
[2]4th Orthopaedic Department, "KAT" General Hospital of Attica, Kifissia, Attica, Greece

Corresponding author:
Sophia Stasi, Laboratory of Neuromuscular and Cardiovascular Study of Motion, Physiotherapy Department, Faculty of Health and Care Sciences, University of West Attica, Campus 1, 28 Agiou Spyridonos St., Egaleo, Attica 12243, Greece.
Email: soniastasi1@gmail.com

Introduction

Osteoarthritis (OA) is among the most prevalent and disabling conditions affecting the elderly and is a particular burden when it affects the hips and knees, as pain and stiffness in these large weight-bearing joints often leads to significant disability requiring surgical intervention.[1,2] Epidemiological studies indicate that hip joint osteoarthritis occurs in 88/100,000 individuals.[3,4] There is an estimated 25% lifetime risk of symptomatic hip OA in people who live to age 85,[2] and an almost 10% lifetime risk of undergoing total hip replacement for end-stage OA.[5] In the Greek population hip OA has a prevalence of 0.9/1000: 1.5/1000 in women and 0.3/1000 in men.[6]

The treatment of hip OA is aimed at functional enhancement and pain management. In an era of evidence-based medicine, a thorough assessment of a patient's pain and functional level requires reliable and validated outcome measures that can provide clinicians with objective and quantified health status data on which to base decisions concerning the most effective treatment plan and surgical versus non-surgical intervention.

Specifically designed outcome measures are frequently used for the evaluation of hip OA patients. The Harris Hip Score (HHS), developed in 1969,[7] is a multidimensional clinician-reported outcome measure that contains 10 items covering pain, function, absence of deformity and range of motion.[8] It is considered an ideal tool for the evaluation of various hip disabilities and methods of conservative treatment, or surgical intervention in adult populations.[9] The HHS has been translated into many languages and its psychometric and measurement properties are well established.[10] In 2000, Byrd and Jones modified the HHS for the long-term evaluation of patients who had undergone hip arthroscopy, deleting the domains concerning deformity and range of motion.[11] This modified HHS (mHHS) includes only assessments based on pain and function; thus, unlike the HHS, it can be used as a patient-reported outcome (PRO) measure. PROs are considered ideal measurement tools for evaluating outcomes because their patient-focused perspective enables patients to actively participate in their own evaluation and to quantify their functional limitations, changes in symptoms over time, and post-treatment outcomes.[12,13] Furthermore, the psychometric properties of PROs are of interest to all practitioners who manage patients, including surgeons, physiotherapists, and researchers.[14]

To our knowledge, the mHHS has been translated and culturally adapted only into Portuguese,[15] while its psychometric properties have not been extensively explored (i.e. there is no information regarding internal consistency for either the original English or the Portuguese version of the mHHS).[16,17] In addition, since the MHHS was designed for the evaluation of patients undergoing hip arthroscopy, there is a lack of information regarding its applicability to patients with chronic hip diseases, such as hip OA.

The purpose of the present study was to evaluate a Greek version of the mHHS as a PRO measurement tool in individuals with hip OA. It was hypothesised that the mHHS would be correlated with the Greek version of other PRO questionnaires and/or objective physical-performance measures (PPM) that are widely used to evaluate hip OA patients. We decided to incorporate PPMs in our study because they have been reported to be good predictors of functionality and they lack the problems inherent in PRO measures, such as patients' inability or unwillingness to answer questions correctly.[18,19]

Methods

Study design and setting

This observational study was conducted in accordance with the ethical principles stated in the Declaration of Helsinki and its later amendments.[20] The Research Committee of the University of West Attica, Athens, Greece and the Scientific Research Council of the "KAT" General Hospital of Attica, Athens, Greece approved the protocol. The study conformed to the "Strengthening the Reporting of Observational studies in Epidemiology" (STROBE) statement for reporting observational studies.[21]

Cultural adaptation of mHHS

Official permission for reprinting and translating the original mHHS questionnaire was given by J. W. Thomas Byrd and Kay S. Jones. Its adaptation into Greek followed the guidelines developed by Guillemin et al.[22,23] and Beaton et al.[24] A team of experts, consisting of health science professionals and two bilingual non-medical specialists, took care of all the required procedures. Technical and linguistic adaptations were agreed in a consensus meeting. Field-testing of the provisional version included its completion by a group of individuals ($n = 20$) from the same target population group, by means of one-to-one interviews, in order to examine the potential distribution of responses and check comprehension. The committee took note of the questions of the participants who completed the provisional version of the questionnaire and proceeded to the cultural adaptation of the final Greek version. First, the sentence 'Please select only one response in each section' was added at the top of the questionnaire. Linguistic adaptations were made in the 'Pain' section with regard to the use of the adjective 'marked' for level of pain. The equivalent term in Greek could be understood as meaning 'remarkable', which was not fully understood by all participants; therefore, another adjective meaning 'intense' was substituted. In addition, in the Distance Walked item of the Function: Gait section, the response option 'Bed and chair' was modified to read 'Moving from bed to chair' and in the Stairs item of the Functional Activities section an addition was made to specify that the word 'Normally' means 'alternating feet on stairs'.

Most of the changes for cross-cultural adaptation were made in the 'Pain' section. This was because most Greeks know pharmaceuticals by their commercial names, rather than by their active ingredient. They are also not used to taking aspirin as an analgesic. Accordingly, references to 'aspirin' in the English questionnaire were supplemented by 'Depon', which is a common brand name of paracetamol that is widely used in Greece. The relevant response option was changed to 'Need to use mild analgesics (Aspirin, Depon)'. In the same section, the English questionnaire refers to 'occasional codeine'. In the final Greek version of the questionnaire, this response option has been changed to 'Occasional need to use strong analgesics (Lonarid)'. Lonarid is an analgesic available in Greece whose main active substance is codeine.

Ultimately, the final version of the modified Harris Hip Score in the Greek language (mHHS-Gr) was completed once again by 2 small groups of individuals ($n = 15$) from the same target population group in order to check comprehension, and to ensure linguistic validation and cross-cultural verification and adaptation. This final version was back-translated into English by a third bilingual non-medical specialist, who did not know the original/English questionnaire. The back-translation was approved by the creators of the original, JW Byrd and KS Jones. The original questionnaire, the Greek-language version and the back-translation of the mHHS-Gr are included in the Supplemental material Appendix.

Participants

110 patients aged 55 years and over, who consulted the co-chief scientific researcher of the present study (GM, orthopaedic surgeon), were invited to participate in this study. The main inclusion criterion was the existence of hip OA according to the Kellgren-Lawrence classification system.[25] Patients who reported pain on active movement of the hip joint and had used anti-inflammatory medication and/or received physical therapy for at least the previous 6 months were eligible for inclusion.[26] Participants were excluded if they: had undergone any kind of surgical intervention to the affected hip; had other hip disorders or medical conditions, such as rheumatoid arthritis, psoriatic arthritis, chronic inflammatory diseases or lower limb muscle weakness due to a central or peripheral neurological aetiology, or were taking medication that adversely affected their postural or dynamic balance. Participants who reported any change in their clinical status or received any treatment interventions between the 2 assessments days were also excluded. Upon acceptance, participants gave their written informed consent and their demographic and clinical characteristics were recorded.

Patient-reported outcomes and physical-performance measures

Modified Harris Hip Score. The mHHS includes assessments based on pain and function. 1 item evaluates pain (0–44 points), while 7 items evaluate the patient's functionality (0–47 points). 3 items – "Limp", "Support" and "Distance Walked", with a score ranging from 0 to 11 – relate to the patient's functionality while walking. The other 4 items evalutate functional activities: "Stairs" and "Socks/Shoes" have a score ranging from 0 to 4, "Sitting" from 0 to 5 and "Public Transportation" from 0 to 1. The total points form a scale from 0 to 91. A multiplier of 1.1 provides a total score of 100 (best possible outcome).[11]

The Greek versions of the PROs Lower Extremity Functional Scale (LEFS-Greek),[27–29] Western Ontario and McMaster Osteoarthritis Index (WOMAC-Gr),[30,31] and the PPMs Timed Up & Go (TUG) test[32,33] and 9-stairs-ascend/desscend (9S-A/D) test[34] were also used in the present study (see Supplemental Material, Appendix).

Procedures

On the initial assessment (day-1), the mHHS-Gr, LEFS-Greek and WOMAC-Gr were given to all participants and completed on site, under the supervision of the same member of the research team. The questionnaires were given out in random order, interspersed with the PPMs (1 questionnaire – TUG test, 2 questionnaires – 9S-A/D test). This allowed sufficient resting time between the tests and reduced the risk of question-order bias. The correct procedures for the TUG and 9S-A/D were carefully explained prior to a single pilot test. Both PPMs were performed only once, so as to minimise habituation bias and avoid affecting the participant's performance. The same researcher recorded all test performance times using a timer with an accuracy of 1/100 second. Participants were allowed to use a walking aid if necessary, but no verbal encouragement or personal assistance was given. Participants were asked to perform the TUG test as quickly as they could while still feeling safe, while for the 9S-A/D test they were asked to proceed in their usual manner, at a safe and comfortable pace, using the stair handrail if necessary. In the present study, the times taken to ascend and descend the stairs were measured separately and the total time was recorded in seconds.

The mHHS-Gr questionnaire was re-administered to all participants 7 days after the first assessment day (day-8).

Patients from our study population who were on the waiting list for a total hip arthroplasty via direct anterior minimal invasive surgery (DAMIS) were used to explore the questionnaire's responsiveness (treatment effect validity).This sub-sample of participants completed the mHHS-Gr questionnaire on 2 further occasions, preoperatively and 4 weeks after DAMIS.

Statistical methods

For the exploration of the psychometric properties of PRO questionnaires, there is a widely-cited rule of thumb that suggests 10 respondents per item.[35,36] The mHHS questionnaire consists of 8 items; thus a sample size of 80 participants would be adequate.

All tests were 2-sided; a *p*-value < 0.05 was considered to denote statistical significance. All analyses were carried out using the statistical package SPSS version 17.00 (Statistical Package for the Social Science, SPSS Inc., Chicago, Ill., USA). Data were expressed as mean ± SD for quantitative variables and as percentages for qualitative variables.

Reliability study

An extensive reliability study was carried out to explore the internal consistency, test-retest reliability (stability) and reproducibility (agreement) of the mHHS-Gr questionnaire. Internal consistency was determined by calculating Cronbach's *alpha* coefficient.[37]

Test-retest reliability (stability) of the mHHS-Gr was determined by calculating the intraclass correlation coefficient (ICC) and its 95% confidence interval (CI).[38] Because this coefficient does not correct for systematic differences and agreement by chance, the scores of the 2 assessments (day-1 and day-8) were tested for systematic differences using the paired *t*-test. Finally, a Bland–Altman plot was used as a visual method of assessing stability.

Reproducibility (agreement) was determined by calculating the floor and ceiling effects and the minimal importance change (MIC). Floor or ceiling effects are considered to be present if more than 15% of respondents achieved the lowest or highest possible score, respectively.[38] The MIC was expressed as $0.5 \times SD$ at baseline. The measurement error is the error of the score not attributed to the construct that is being measured and is expressed as the standard error of measurement (SEM), using the formula $SEM = SD \times \sqrt{(1-ICC)}$, where SD is the standard deviation of all patients at baseline. Minimal detectable change (MDC) is the change of score that exceeds the SEM and was calculated as $SEM \times 1.96 \times \sqrt{2}$ at individual level.[39]

Validity study

Construct validity was defined as the degree to which an outcome score is consistent with another relevant score.[12] The construct validity of MHHA-Gr was determined by its correlations with LEFS-Gr total score, WOMAC-Gr's (total and subscale scores), TUG test and 9S-A/D test performance times, using Pearson's and Spearman's correlation coefficients. The strength of correlation between mHHS-Gr and the well-established PROs and objective PPMs would support the validity of the mHHS-Gr questionnaire in measuring important aspects of functional status in hip OA patients.

The known-groups validity of mHHS-Gr was examined in terms of its ability to distinguish between subgroups of patients formed on the basis of their functional status according to TUG test performance time (cut-off value 13.5 seconds). The TUG test was chosen as an external criterion because it has previously been used as an indicator for discriminating hip OA patients who underwent total arthroplasty according to their poor (>13.5 seconds) or good (<13.5 seconds) functional status.[40] An independent samples *t*-test was used for the statistical analysis.

Receiver operating curve (ROC) analysis was conducted to determine the cut-off point of mHHS-Gr total score for differentiation between subgroups of patients formed on the basis of their functionality. The area under the curve (AUC), standard error and 95% CI were calculated using the maximum likelihood estimation method, and the sensitivity and specificity of different cut-off points for mHHS-Gr as a measure of functional status were estimated using TUG test performance time (cut-off value of 13.5 seconds) as an estimated variable.

Responsiveness

Responsiveness (treatment effect validity) was examined in terms of the questionnaire's ability to monitor changes 4 weeks after DAMIS-THA surgery. The preoperative and postoperative mHHS-Gr score differences were evaluated by calculating the standardised response mean (SRM) using the formula

$$SRM = Mean_{Postoperative} - Mean_{Preoperative} / Standard\ deviation_{Postoperative - Preoperative}$$

and Guyatt's responsiveness statistic (GRS) using the formula

$$GRS = MCID \ / \ \sqrt{2} \times MSE_{mHHS\text{-}Gr}.$$

MCID is the minimal clinically important difference observed in our sample and $MSE_{mHHS\text{-}Gr}$ represents the mean square error of MHHS-Gr obtained from the analysis of variance (ANOVA).[41]

Results

Descriptive and clinical data

In order to explore the reliability and validity properties of the questionnaire, 110 patients were assessed for eligibility from February until July 2017. The data from 90 participants were analysed (Figure 1). Regarding mHHS-Gr responsiveness after treatment, data from a sub-sample ($n = 30$) of our participants who were suffering from a late

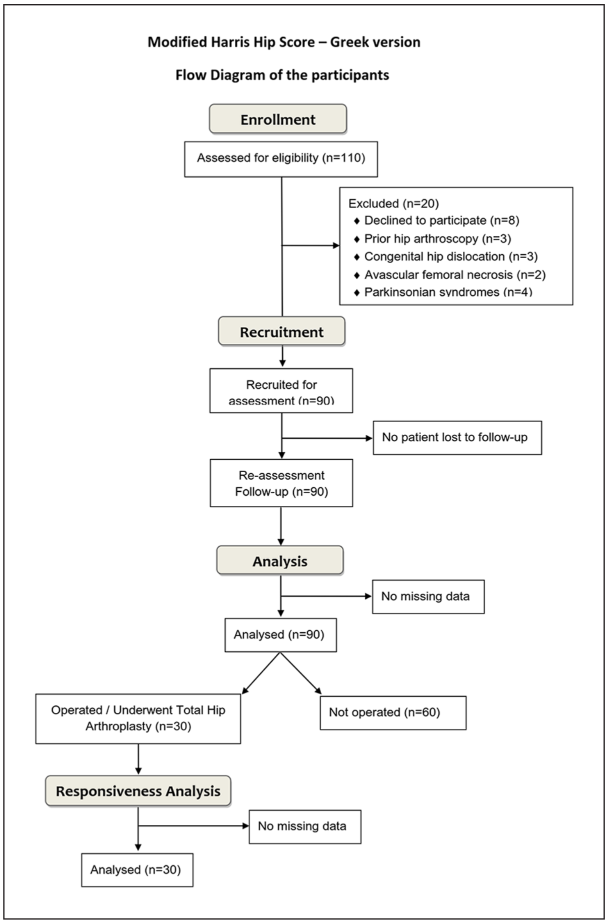


Figure 1. Flow diagram of the enrolment of participants.

stage of hip OA and underwent DAMIS-THA were analysed. This phase of the study lasted from September 2017 until January 2018 (Figure 1). The demographic characteristics and clinical measurements of all the participants are shown in Table 1.

Reliability properties

The internal consistency of mHHS-Gr was measured by Cronbach's *alpha* and estimated as 0.614, which indicates moderate internal consistency (Table 2). The most important item of the scale was the "Function: Gait–Limp" based on Cronbach's *alpha* if the item was deleted (0.472) (Table 2). As regards test–retest reliability, the paired samples *t*-test between initial assessment and reassessment of mHHS-Gr total score indicated no statistically significant difference ($p=0.277$). ICC (95% CI) between initial assessment and reassessment of the mHHS-Gr total was 0.94 (0.91–0.97) ($p<0.001$). Bland-Altman plots showed that almost all differences were within mean difference ± 2 SDs, confirming agreement between the 2 assessments (Figure 2). These results indicate that mHHS-Gr scores were remarkably consistent between the two measurements. As regards reproducibility (agreement), the floor and ceiling effects were 1.1% and 1.1%, respectively. The critical value of 15% was not exceeded, so there was neither a ceiling nor a floor effect for mHHS-Gr. SEM, MIC and MDC values are presented in Table 2.

Validity properties

Table 3 summarises the correlation between the mHHS-Gr and the selected validation instruments. The correlation coefficient ranged from −0.533 to −0.786 ($p<0.001$) for negative correlations and was 0.801 for positive correlations ($p<0.001$). The negative correlations can be explained by the fact that higher scoring in mHHS implies better functional status, whereas higher scoring in WOMAC and longer performance time in the TUG and 9S-A/D tests are equivalent to poorer functional status. The positive correlations can be explained by the fact that higher scoring in mHHS and in LEFS implies better functional status. The above results indicate that mHHS-Gr was significantly associated with all validation criteria, confirming the instrument's construct validity.

The analysis of known-groups validity showed that the mHHS-Gr total score was statistically significantly higher in participants with a TUG score <13.5 than in those with TUG time >13.5 ($p<0.001$) (Table 3). In ROC analysis the AUC for mHHS-Gr total score was 0.775 (95% CI, 0.68–0.87; $p<0.001$), with cut-off point 52.5, and sensitivity 81% and specificity 71% (Figure 3).

Responsiveness properties

The preoperative mHHS-Gr average score was 35.19 ± 7.73, while the 4th postoperative week's average score was 60.50 ± 6.84. The paired-sample test difference [MCID (95% CI)] between the preoperative and 4th postoperative week's values of mHHS-Gr was 25.31 (23.86–26.76) ($p<0.0005$). The responsiveness after treatment of mHHS-Gr was measured by determining the SRM and GRS (Table 2). The values of both indices were greater than 0.8,[41] indicating that mHHS-Gr is effective for detecting treatment changes in patients with hip OA.

Discussion

Worldwide, this is the first study to examine the internal consistency of the MHHS at the same time as other reliability properties (test-retest reliability, and reproducibility), in a sample consisting solely of patients with hip OA This is also the first study to use both PROs and PPMs to examine the validity properties of the mHHS in hip OA patients, and to examine the questionnaire's responsiveness after DAMIS. The mHHS-Gr was found to have excellent reliability properties, exhibited significant validity against the

Table 1. Demographic characteristics and clinical measurements of the study sample ($n = 90$).

Characteristics/measurements	Values	Range
Age (years)[a]	66.28 ± 8.27	55–87
Height (metres)[a]	1.66 ± 0.085	1.48–1.88
Weight (kg)[a]	77.86 ± 14.90	55–126
Body mass index (kg/m^2)[a]	28.13 ± 4.36	20.28–39.18
Sex: men/women (%)	26.7/73.3	
Dominant lower limb: right/left (%)	86.7/13.3	
Affected hip: right/left (%)	48/42	
Kellgren & Lawrence classification (%)		
Grade 1/Grade 2/Grade 3/Grade 4	1.1/15.6/47.8/35.6	
Use of walking aid (%)	24.4	
Modified Harris Hip Score-Greek version*[a]	52.14 ± 15.55	20.9–95.7
Lower Extremity Functional Scale-Greek version[a]	27.28 ± 27.28	6–74
WOMAC LK 3.1 Greek for Greece Index–Pain[a]	4.54 ± 1.99	0–8.5
WOMAC LK 3.1 Greek for Greece Index–Stiffness[a]	5.18 ± 2.49	0–10
WOMAC LK 3.1 Greek for Greece Index–Function[a]	5.35 ± 2.019	0.29–9.26
WOMAC LK 3.1 Greek for Greece Index–Total score[a]	15.07 ± 2.2	1.54–27.76
Timed Up and Go test performance time (s)[a]	15.21 ± 5.37	6.79–43.86
9 stairs-ascend/descend test performance time (s)[a]	20.21 ± 10.40	8.18–50.01

kg, kilograms; s, seconds; WOMAC, Western Ontario and McMaster University index.
[a]The values are expressed as mean ± standard deviation.
*The initial assessment's values of modified Harris Hip Score - Greek version were presented.

Table 2. Reliability properties ($n = 90$) and responsiveness ($n = 30$) of modified Harris Hip Score-Gr.

	Items		Cronbach's alpha "if item deleted"
Internal consistency	Pain		0.645
	Function: Gait–Limp		0.472
	Function: Gait–Support		0.600
	Function: Gait–Distance Walked		0.557
	Functional Activities: Stairs		0.598
	Functional Activities: Socks/Shoes		0.592
	Functional Activities: Sitting		0.586
	Functional Activities: Public Transportation		0.614
	Overall Cronbach's *alpha*		0.614
Test-retest reliability	ICC 95% CI	0.948 (0.91–0.97)	$p < 0.001$
	Paired samples *t*-test	51.49 ± 16.3 & 50.70 ± 16.15[a]	NS (0.277)
	SEM	3.54	
Reproducibility	MDC	10.39	MIC < MDC
	MIC	7.75	
Responsiveness	SRM	6.48	
	MCID (95% CI)	25.31 (23.86–26.76)	$p < 0.0005$
	GRS	2.4	

ICC, intraclass correlation coefficient; CI, confidence interval; SEM, standard error of measurement; MDC, minimal detectable change; MIC, minimal important change; SRM, standardised response mean; MCID, minimal clinically important difference; GRS, Guyatt's responsiveness statistic.
[a]The values of modified Harris Hip Score-Gr at initial assessment and reassessment expressed as mean ± standard deviation.

LEFS-Greek, WOMAC-Gr, TUG and 9S-A/D tests, and showed excellent ability to detect treatment effects.

Reliability

Our results indicate that the mHHS-Gr has high reliability properties and may be used in clinical practice and research to evaluate Greek hip OA patients. To our knowledge, there is no information regarding internal consistency for either the original/English or the Portuguese versions of the mHHS.[15–17] Analysis of the internal consistency of mHHS-Gr showed that the eight items are moderately interdependent and homogeneous in terms of the construct they measure. Calculation of "Cronbach's *a* if item deleted" revealed that the most important item was the "Function: Gait–Limp". This may be explained by the fact that 83.4% of our sample consisted of patients with grade 3 or 4 hip OA. It has been reported that in advanced stages of the disease, like most of our participants, the most common symptoms are pain and limited active and passive hip motion, which affects hip mobility and leads to limping and disturbed walking dynamics.[42]

Test-retest reliability indicates the stability of an individual's response over time.[38] The mHHS-Gr total scores revealed no statistically significant difference ($p = 0.277$) between the 2 assessments (day-1 and day-8), indicating excellent stability of the patient's response over time.

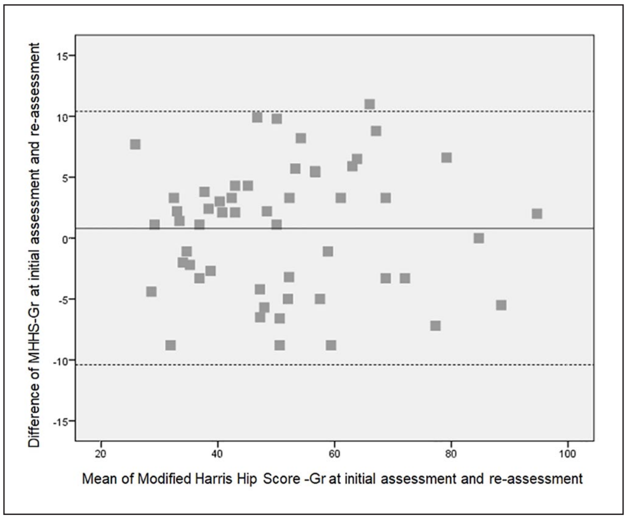


Figure 2. Bland-Altman plot of modified Harris Hip Score - Greek version, showing difference in means between initial assessment and reassessment: 0.79 (95% CI, −0.65 to +2.24).

Moreover, mHHS-Gr's ICC values were above the level of 0.90 ($p < 0.001$) (38). It has been reported that a PRO may be deemed adequate for use in groups (research) if the ICC is >0.8 and for use in patients (clinical practice) if the ICC is >0.9.[43,44]

Reproducibility (agreement) refers to the degree to which one can assign qualitative meaning to quantitative scores.[45] There was neither a ceiling nor a floor effect in mHHS-Gr total scores. The SEM was less than four points compared with a total score of 100 points. The mHHS-Gr MIC was smaller than its MDC; this indicated that the agreement of mHHS is satisfactory and scored positive by the criteria of Terwee et al.[38]

Although the reliability properties of the original/English questionnaire of the MHHS were not explored by Byrd and Jones,[11] there are studies involving patients who had undergone hip arthroscopic surgery,[14] or young adults with femoroacetabular impingement,[46] in which mHHS test-retest ICC values were reported. In the study of Kemp et al.[14] the ICC value was 0.91, consistent with our ICC value, but in the study of Hinman et al.[46] the ICC value was 0.76. This lower ICC value might be explained by the fact that participants did not complete the MHHS questionnaire again in the second session of this study: instead they completed a seven-point Likert Global Perceived Effect scale to determine if any substantial change in their condition had occurred over the interval between test sessions.[46]

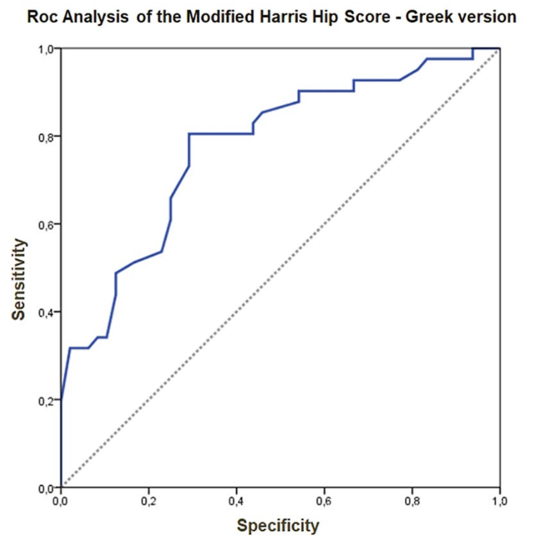
Validity

Our validity results showed that the mHHS is adequate for use in hip OA patients. mHHS-Gr was significantly associated with all selected validation criteria, presenting high

Table 3. Validation properties of the modified Harris Hip Score - Greek version.

Construct validity			
Validation instruments		Modified Harris Hip Score - Greek version	*p*-value
Lower Extremity Functional Scale – Greek version		0.801[a]	<0.001
WOMAC® LK 3.1 Greek for Greece Index–Total		–0.783[a]	<0.001
WOMAC® LK 3.1 Greek for Greece Index–Pain		–0.728[a]	<0.001
WOMAC® LK 3.1 Greek for Greece Index–Stiffness		–0.593[a]	<0.001
WOMAC® LK 3.1 Greek for Greece Index–Function		–0.786[a]	<0.001
Timed Up and Go Test		–0.547[b]	<0.001
9stairs-ascend/descend Test		–0.575[b]	<0.001
Known-groups validity			
Subgroups of patients[c]	*n*	Mean ± SD[d]	*p*-value
TUG performance time less than 13.5 s[c]	48	59.00 ± 14.16	<0.001
TUG performance time more than 13.5 s[c]	41	44.62 ± 13.32	

[a]Pearson's correlation coefficient.
[b]Spearman's correlation coefficient.
[c]Timed Up and Go performance time (cut-off 13.5 s) as estimated variable.
[d]Mean ± SD of modified Harris Hip Score - Greek version at initial assessment.



AUC	*p*-value	Cut-off Point	Sensitivity	Specificity	95%CI
0.775	<0.005	52.5	81%	71%	0.677 0.872

Figure 3. Receiver operating curve (ROC) analysis of the modified Harris Hip Score - Greek version using the TUG score (cut-off 13.5 seconds) as estimated variable.

correlations with the other PROs, whereas correlations with the PPMs were moderate. Regarding PROs, the strongest correlations were observed between mHHS-Gr, LEFS-Gr and the WOMAC-Gr outcomes – apart from the WOMAC-Gr Stiffness subscale. This moderate association may be explained by the fact that the subscale was designed to examine the stiffness experienced by the patient after first waking in the morning and later in the day after rest or sitting, not during the functional activities described in MHHS items. Correlations with the TUG test and the 9S-A/D test were also moderate, but this is not surprising because PPMs and PROs assess different dimensions of functionality (patient ability vs. patient perception).[47] However, the 2 assessment methods provide complementary information,[48] and both are needed to perceive the multidimensional impact of functionality in its entirety,[47] which is essential to clinical research and practice involving hip OA patients.

In 2 studies involving hip arthroscopy populations, the construct validity of the mHHs was investigated against the Medical Outcomes Study 36-Item Short-Form Health Survey (SF-36). In the study of Kemp et al.[14], the Pearson's *r* correlation coefficients for SF-36–Bodily Pain, SF-36–Physical functioning and SF-36–Physical Role were 0.604, 0.703 and 0.480, respectively ($p=0.05$). In the study of Potter et al.[49] Pearson's *r* for the SF-36–Bodily pain, SF-36–Physical functioning and the SF-36–Physical Component subscale scores were 0.73, 0.71 and 0.85, respectively ($p<0.001$). Even though these studies involved hip arthroscopy populations, the values of the correlation coefficients are in line with our results. The mHHS appears to be a hip-specific PRO measure that can be used in young-to-middle aged as well as elderly groups of patients who exhibit various hip disorders and is compatible with generic (SF-36), disease-specific (WOMAC) or functional-specific (LEFS) PROs.

Known-group analysis of the data showed that mHHS-Gr could detect statistically significant differences in total scores between subgroups of patients, based on their functional status according to TUG test performance time. This is also the first study of MHHS to determine specific cut-off points for functional status. Our results indicate that hip OA patients with an MHHS-Gr total score < 52.5 have an 81% probability of having a TUG test performance time <13.5, while patients with MHHS-Gr >52.5 have a 71% probability of having a TUG test performance time >13.5. Therefore, an MHHS-Gr score of 52.5 may be used as a cut-off value for the determination of the functional status of hip OA patients with similar characteristics to our study group.

Responsiveness

Assessment of sensitivity to change is important if a PRO is to be used in treatment evaluation studies. The large magnitude of SRM and GRS in this study provides evidence that mHHS-Gr is a responsive assessment with excellent ability to detect treatment effects. The SRM value in this study is higher than that reported by previous studies, which reported moderate responsiveness (SRM=0.588) of mHHS in younger patients who had undergone hip arthroscopy.[14] This difference in SRM values may be explained by the fact that patients of our responsiveness sub-sample were suffering from late stages of hip OA and had low mHHS-Gr preoperative scores (35.19 ± 7.73) compared with the hip arthroscopy patients (86.6 ± 14.6).[14] It has been reported that severely affected patients are more likely to experience larger improvements from major surgical treatment.[50] Overall, the mHHS is a sensitive PRO measure for detecting changes in patient's pain and function after treatment.

Strengths and limitations

The restrictive inclusion and exclusion criteria for the participant's selection from a well-defined target population are an important strength of this study. In addition, our sample size of 90 participants meets the guidelines for the exploration of psychometric properties of PRO questionnaires.[35,36] Moreover, conducting an extensive reliability study and examining the validity properties of the mHHS against both PROs and PPMs, as well as exploring the questionnaire's responsiveness, added statistical power to our results.

However, the study also has some potential limitations. First, since the Harris Hip Score questionnaire has not been officially translated into the Greek language, we are unable to examine its correlations with mHHS-Gr. Furthermore, we did not examine certain mHHS-Gr measurement properties, such as the content or face validity. Finally, the study design did not include factor analysis of the questionnaire.

Conclusion

The results shown here indicate that the mHHS-Gr questionnaire has high reliability properties, presenting strong correlations with other PROs and satisfactory correlations with PPMs, and showed excellent responsiveness in the detection of treatment effects. The mHHS could possibly be used as a PRO in clinical practice and research to evaluate hip OA patients. Further research is needed in order to confirm our results and to explore the questionnaire's reliability properties in different groups of patients, its validity properties against other PROs, and its responsiveness after treatments other than DAMIS. A broader awareness of the findings in the Greek setting would facilitate objective comparisons between studies with different national origins and could contribute to the validity of mHHS in future meta-analyses.

Acknowledgements

The authors would like to thank Dr Antonios Galanos, Biostatistician of Laboratory of Research of the Musculoskeletal System (LRMS), Faculty of Medicine, National and Kapodistrian University of Athens, Greece, for the statistical analysis of the data and Mr Philip Lees, medical writer, for his invaluable editorial assistance with the English text.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship and/or publication of this article.

Supplemental material

Supplemental material for this article is available online.

References

1. Palazzo C, Nguyen C, Lefevre-Colau MM, et al. Risk factors and burden of osteoarthritis. *Ann Phys Rehabil Med* 2016; 59: 134–138.
2. Murphy LB, Helmick CG, Schwartz TA, et al. One in four people may develop symptomatic hip osteoarthritis in his or her lifetime. *Osteoarthritis Cartilage* 2010; 18: 1372–1379.
3. Culliford DJ, Maskell J, Kiran A, et al. The lifetime risk of total hip and knee arthroplasty: results from the UK general practice research database. *Osteoarthritis Cartilage* 2012; 20: 519–524.
4. Oliveria SA, Felson DT, Reed JI, et al. Incidence of symptomatic hand, hip, and knee osteoarthritis among patients in a health maintenance organization. *Arthritis Rheum* 1995; 38: 1134–1141.
5. van Saase JL, van Romunde LK, Cats A, et al. Epidemiology of osteoarthritis: Zoetermeer survey. Comparison of radiological osteoarthritis in a Dutch population with that in 10 other populations. *Ann Rheum Dis* 1989; 48: 271–280.
6. Andrianakos AA, Kontelis LK, Karamitsos DG, et al. Prevalence of symptomatic knee, hand, and hip osteoarthritis in Greece. The ESORDIG study. *J Rheumatol* 2006; 33: 2507–2513.
7. Harris WH. Traumatic arthritis of the hip after dislocation and acetabular fractures: treatment by mold arthroplasty. An end-result study using a new method of result evaluation. *J Bone Joint Surg Am* 1969; 51: 737–755.
8. Marchetti P, Binazzi R, Vaccari V, et al. Long-term results with cementless Fitek (or Fitmore) cups. *J Arthroplasty* 2005; 20: 730–737.
9. Nilsdotter A and Bremander A. Measures of hip function and symptoms: Harris hip score (HHS), hip disability and osteoarthritis outcome score (HOOS), Oxford hip score (OHS), Lequesne index of severity for osteoarthritis of the hip (LISOH), and American academy of orthopedic surgeons (AAOS) hip and knee questionnaire. *Arthritis Care Res (Hoboken)* 2011; 63(Suppl. 11): S200–S207.
10. Munro C and Johnston AT. Outcome measures following hip arthroplasty. *J Orthop Trauma* 2018; 32: 34–37.
11. Byrd TJ and Jones KS. Prospective analysis of hip arthroscopy with 2-year follow-up. *Arthroscopy* 2000; 16: 578–587.
12. Mokkink LB, Terwee CB, Patrick DL, et al. 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* 2010; 19: 539–549.
13. Institute of Medicine. *Crossing the quality chasm: a new health system for the 21st century*. Washington, DC: National Academy Press, 2001.
14. Kemp JL, Collins NJ, Roos EM, et al. Psychometric properties of patient-reported outcome measures for hip arthroscopic surgery. *Am J Sports Med* 2013; 41: 2065–2073.
15. Guimarães RP, Alves DP, Azuaga TL, et al. Translation and transcultural adaptation of the modified Harris hip score. *Acta Ortopédica Brasileira* 2010; 18: 339–342.
16. Tijssen M, van Cingel R, van Melick N, et al. Patient-reported outcome questionnaires for hip arthroscopy: a systematic review of the psychometric evidence. *BMC Musculoskelet Disord* 2011; 12: 117.
17. Ramisetty N, Kwon Y and Mohtadi N. Patient-reported outcome measures for hip preservation surgery-a systematic review of the literature. *J Hip Preserv Surg* 2015; 2: 15–27.
18. Juhakoski R, Tenhonen S, Anttonen T, et al. Factors affecting self-reported pain and physical function in patients with hip osteoarthritis. *Arch Phys Med Rehabil* 2008; 89: 1066–1073.
19. Rockwood K, Awalt E, Carver D, et al. Feasibility and measurement properties of the functional reach and the timed up and go tests in the Canadian study of health and aging. *J Gerontol A Biol Sci Med Sci* 2000; 55: M70–M73.

20. World Medical Association. WMA declaration of Helsinki – ethical principles for medical research involving human subjects, https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/ (2017, accessed 13 October 2017).
21. von Elm E, Altman DG, Egger M, et al. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet* 2007; 370: 1453–1457.
22. Guillemin F, Bombardier C and Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *J Clin Epidemiol* 1993; 46: 1417–1432.
23. Guillemin F. Cross-cultural adaptation and validation of health status measures. *Scand J Rheumatol* 1995; 24: 61–63.
24. Beaton DE, Bombardier C, Guillemin F, et al. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine (Phila Pa 1976)* 2000; 25: 3186–3191.
25. Kellgren JH and Lawrence JS. Radiological assessment of osteo-arthrosis. *Ann Rheum Dis* 1957; 16: 494–502.
26. Puopolo A, Boice J, Fidelholtz J, et al. A randomized placebo-controlled trial comparing the efficacy of etoricoxib 30 mg and ibuprofen 2400 mg for the treatment of patients with osteoarthritis. *Osteoarthritis Cartilage* 2007; 15: 1348–1356.
27. Binkley JM, Stratford PW, Lott SA, et al. The lower extremity functional scale (LEFS): scale development, measurement properties, and clinical application. North American orthopaedic rehabilitation research network. *Phys Ther* 1999; 79: 371–383.
28. Stasi S, Papathanasiou G, Anagnostou M, et al. Lower extremity functional scale (LEFS): cross – cultural adaption into Greek and reliability properties of the instrument. *Health Sci J* 2012; 6: 750–773.
29. Stasi S, Papathanasiou G, Korres N, et al. Validation of the lower extremity functional scale in community-dwelling elderly people (LEFS-Greek); determination of functional status cut-off points using TUG test. *Eur Geriatr Med* 2013; 4: 237–241.
30. Bellamy N and Buchanan WW. A preliminary evaluation of the dimensionality and clinical importance of pain and disability in osteoarthritis of the hip and knee. *Clin Rheumatol* 1986; 5: 231–241.
31. Papathanasiou G, Stasi S, Oikonomou L, et al. Clinimetric properties of WOMAC index in Greek knee osteoarthritis patients: comparisons with both self-reported and physical performance measures. *Rheumatol Int* 2015; 35: 115–123.
32. Podsiadlo D and Richardson S. The timed "Up & Go": a test of basic functional mobility for frail elderly persons. *J Am Geriatr Soc* 1991; 39: 142–148.
33. Herman T, Giladi N and Hausdorff JM. Properties of the 'timed up and go' test: more than meets the eye. *Gerontology* 2011; 57: 203–210.
34. Nightingale EJ, Pourkazemi F and Hiller CE. Systematic review of timed stair tests. *J Rehabil Res Dev* 2014; 51: 335–350.
35. Nunnally J. *Psychometric methods*. New York: McGraw-Hill, 1978, p. 421.
36. Cappelleri JC, Lundy JJ and Hays RD. Overview of classical test theory and item response theory for quantitative assessment of items in developing patient-reported outcome measures. *Clin Ther* 2014; 36: 648–662.
37. Nunnally JC and Bernstein IH. *Psychometric theory*. 3rd ed. New York: McGraw-Hill, 1994.
38. Terwee CB, Bot S, de Boer M, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007; 60: 34–42.
39. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res* 2005; 19: 231–240.
40. Kamimura A, Sakakima H, Tsutsumi F, et al. Preoperative predictors of ambulation ability at different time points after total hip arthroplasty in patients with osteoarthritis. *Rehabil Res Pract* 2014; 2014: 861268.
41. Luiz RR and Almeida RMVR. On the measurement of change in medical research. *Int J Stat Med Res* 2012; 1: 144–147.
42. Jotanovic Z, Mihelic R, Gulan G, et al. Osteoarthritis of the hip: an overview. *Periodicum Biologorum* 2015; 117: 95–108.
43. Roos EM, Engelhart L, Ranstam J, et al. ICRS recommendation document: patient-reported outcome instruments for use in patients with articular cartilage defects. *Cartilage* 2011; 2: 122–136.
44. Terwee CB, Dekker FW, Wiersinga WM, et al. On assessing responsiveness of health-related quality of life instruments: guidelines for instrument evaluation. *Qual Life Res* 2003; 12: 349–362.
45. Lohr KN, Aaronson NK, Alonso J, et al. Evaluating quality of life and health status instruments: development of scientific review criteria. *Clin Ther* 1996; 18: 979–992.
46. Hinman RS, Dobson F, Takla A, et al. Which is the most useful patient-reported outcome in femoroacetabular impingement? Test-retest reliability of six questionnaires. *Br J Sports Med* 2014; 48: 458–463.
47. Wright AA, Cook CE, Baxter GD, et al. Relationship between the Western Ontario and McMaster universities osteoarthritis index physical function subscale and physical performance measures in patients with hip osteoarthritis. *Arch Phys Med Rehabil* 2010; 91: 1558–1564.
48. Stratford PW and Kennedy DM. Performance measures were necessary to obtain a complete picture of osteoarthritic patients. *J Clin Epidemiol* 2006; 59: 160–167.
49. Potter BK, Freedman BA, Andersen RC, et al. Correlation of short form-36 and disability status with outcomes of arthroscopic acetabular labral debridement. *Am J Sports Med* 2005; 33: 864–870.
50. Escobar A, Pérez LG, Herrera-Espiñeira C, et al. Total knee replacement; minimal clinically important differences and responders. *Osteoarthritis Cartilage* 2013; 21: 2006–2012.