Psychometric Assessment of the Hip Disability and Osteoarthritis Outcome Score: Measuring Disability and Function Following Total Hip Arthroplasty

A Dissertation
Presented in Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy
with a
Major in Education, Health and Human Sciences
in the
College of Graduate Studies
University of Idaho
by

Emilie N. Miley

Major Professor: Russell T. Baker, Ph.D., DAT
Committee Members: Scott W. Cheatham, Ph.D., DPT; Lindsay W. Larkins, DAT;
Michael A. Pickering, Ph.D.
Department Administrator: Philip W. Scruggs, Ph.D.

December 2023

Abstract

Evidence-based practice (EBP) in healthcare is imperative for optimum patient care. As part of the EBP process, clinicians must collect, appraise, and interpret clinical research to help guide their clinical practice. To help support EBP in clinical practice, patient-reported outcome (PRO) measures may be collected. Clinicians use PRO measures to assess the outcomes related to different health-related dimensions (e.g., pain, function, quality of life) spanning across several different intervention types (e.g., rehabilitation, surgical intervention). Not only do clinicians collect and interpret these PROs in a single setting, but also assess these outcomes over numerous visits (e.g., preoperative, postoperative).

As PROs are important for EBP, focus now exists on the collection of these outcomes for purposes of reimbursement across healthcare facilities nationwide. More specifically, collecting outcomes associated with surgical interventions such as total hip arthroplasty (THA) are imperative as the number of procedures expected to rise substantially over the next decade. However, many PROs related to THA have minimally been tested using recommended contemporary psychometric analysis techniques. Therefore, the purpose of this dissertation was to assess the psychometric properties of two outcomes used in clinical practice to assess hip-related interventions such as THA: 1.) Hip Disability and Osteoarthritis Outcome Score Joint Replacement (HOOS-JR) and 2.) Hip Disability and Osteoarthritis Outcome Score (HOOS).

The HOOS-JR is a one-factor 6-item instrument designed to assess outcomes associated pain and function following a THA. The HOOS-JR was originally derived as a short-form from the original 40-item HOOS; however, minimal research has been identified pertaining to the assessment of the scale structure in a group of patients undergoing THA. Therefore, a confirmatory factor analysis (CFA) was conducted on the HOOS-JR to determine if the scale met the recommended goodness-of-fit indices in a large population of patients who underwent a THA. The HOOS-JR underwent further refinement due to the lack of model fit, which yielded a one-factor 5-item HOOS-JR. The alternate 5-item HOOS-JR was then subjected to invariance testing between age groups and sex (i.e., multi-group), and longitudinally across five time points (i.e., preoperatively and 6-months, 1-year, 2-years, and 3-years postoperatively). In addition, latent growth curve (LGC) modeling was performed to determine intraindividual (i.e., differences in mean scores over time) and interindividual (i.e., differences in mean scores between groups) differences in the responses to the outcome over time. In addition, LGC modeling was assessed to determine if the rate of change in these scores were linear. The 5-item HOOS-JR slightly exceeded the recommended cut-off values for testing; upon further assessment, item five demonstrated item-level bias within the model. In addition, no differences in mean scores were found between groups, and the change in mean scores were not consistent with a linear trajectory. These findings suggest a more curvilinear model was identified with most of the improvement in mean scores to occur between preoperative and 6-months postoperative. In its current form, the 6-item HOOS-JR is not recommended for use in clinical practice. However, the 5-item HOOS-JR maybe suitable though caution is warranted when attempting to make comparisons in mean scores across groups or over time. Continued refinement is recommended before implementation of the modified 5-item HOOS-JR in clinical practice and research.

The 40-item HOOS was created as a region- and disease-specific outcome that assesses five dimensions of hip-related health, including pain, symptoms, functional limitations, function associated with higher level activities, and hip related quality of life. Previous research conducted revealed the HOOS scale structure did not meet the recommended goodness-of-fit indices in a sample of mostly healthy individuals. In addition, analyses of the scale structure had not been confirmed in a large sample of patients who underwent a THA. Therefore, a CFA was conducted on the five-factor, 40-item HOOS. As model fit indices were not met, the scale was split into two samples (n1 and n2). An exploratory factor analysis (EFA) was performed on sample n1 to identify if a more parsimonious structure (i.e., alternate HOOS-9) could be found. Once an alternate model was found, CFA procedures were conducted on the sample n2. Recommended goodness-of-fit indices were met for the alternate model; therefore, the two samples were combined back together. The scale was then subjected to invariance testing between age groups and sex (i.e., multi-group) and longitudinally over five time points (i.e., preoperative and 6-months, 1-year, 2-years, and 3-years postoperatively). In addition, LGC modeling was conducted to determine if there were differences (i.e., intraindividual and interindividual) in mean scores between time points or between groups. Also, LGC modeling was used to determine if improvement in scores were linear. Upon our results, the alternate three-factor HOOS-9 met recommended goodness-of-fit indices for both groups, therefore comparisons in mean scores could be assessed across age groups and sex (i.e., males and females). Recommended model fit indices were also met at each time point and across time, suggesting the mean scores of the alternate HOOS-9 can be compared across time points. Lastly, the LGC model demonstrated the change in scores were not linear, however, most improvement in mean scores occurred between preoperatively to 6-months postoperatively. Additionally, there were significant changes in the mean scores across patients between preoperative to 3-years postoperatively, however, no significant differences were noted between groups. Therefore, in its current form, the 40-item HOOS should not be used in clinical practice and research; however, the alternate HOOS-9 may be a more viable option.

Acknowledgments

To Dr. Russell Baker, I am truly humbled by the opportunity you provided me to pursue my Ph.D. I would not have considered continuing my education further if it were not for our many conversations. I am genuinely grateful for the mentorship, friendship and support you demonstrated during my time at the University of Idaho, and I could not have completed this journey without you guiding me along the way. Thank you for encouraging me to step out of my comfort zone and providing a safe environment to grow. Thank you for pushing me to alter my dissertation topic even though I did not want to. You allowed me to lean on you even when things got tough. Thank you for challenging me to be the best version of myself. Thank you for giving me the strength and the support I needed to become the scholar I am today.

Dr. Lindsay Larkins, without running into you at the CAATE meeting, I would have never considered pursuing my Ph.D. You planted the seed to consider the opportunity to apply, and I am truly thankful for your kind words of encouragement. The compassion and enthusiasm you demonstrated along the way helped push me across the finish line. Thank you for allowing me to lean on you along this journey.

Thank you, Dr. Tony Pickering, for teaching the course that steered me towards the line of research I pursue today. I learned so much from the passion you demonstrated and our intellectual conversations which constantly provided me with so much learning to absorb. You challenged me and provided a safe learning environment so I could take the feedback and continue to grow and excel.

Thank you, Dr. Scott Cheatham, for the passion you bestowed onto me through my time in the DAT related to one of the most complicated regions of the body. Without your knowledge and commitment to education, my line of research may be very different. I am thankful for the conversations we have had over the years, and I am truly grateful for the inspiration you provided me to continue to grow as a clinician-scholar.

Dr. Madeline Casanova, words cannot express enough how grateful I am to have met you along this journey. You have brought me so many stimulating conversations and rallied for me over the years. Your support has meant so much to me and I am extremely lucky to have you as my friend and colleague. I look forward to our long conversations and continued collaboration in the future.

To many of my colleagues along the way, especially Drs. Jayme Baker, Jim May, Alan Nasypany and Chancellor Gray. Thank you for the continued support and guidance. I am forever grateful to have had the experience working with you all. Each one of you played a special role in my journey.

Dedication

To my parents.

You taught me to not settle, reach big, and for whatever I was doing, give it my all. Thank you for your love and support.

To Joe and Ashlee.

Thank you for influencing me to be the person I am today. Thank you for the strength you demonstrated to me along the years. I am grateful to be able to be called your little sister.

To Megan and Rodrigo.

My best friends, my confidants. Thank you for helping me survive this journey. We have shared so many memories exploring the country and I cannot wait to make more. I am truly blessed with you both in my life and grateful to call you my best friends.

To Norma, Mary, and Catalina.

You all supported me throughout this journey. We have endured so many great times together, and I am blessed to call you, my friends.

To Bentley and Ella.

You have been my constant. Your love for the outdoors and happiness to travel with me anywhere I wanted to go has allow us to have so many adventures together. You have kept me on my toes and pushed me to get outdoors even when I did not want to. Thank you for your endless love.

To Josh.

Words cannot express my appreciation for you in my life. You were my rock and my inspiration. I am so grateful for all the support you provided me, and I could not have accomplished this without you by my side.

Table of Contents

List of Tables

List of Figures

Statement of Contribution

Dr. Emilie Miley confirms the following contributions to this study: study conception, design, statistical analyses, interpretation of results, and manuscript preparation. In addition, Dr. Russell Baker attributed to the following: reviewing the study conception, assisted in gaining access to data, obtaining institutional review board approval, reviewed methods and results, provided mentorship during statistical analyses, and contributed manuscript preparation. Dr. Scott Cheatham and Dr. Lindsay Larkins contributed to manuscript preparation. Dr. Tony Pickering and Dr. Madeline Casanova provided mentorship during statistical analyses and contributed to manuscript preparation. All authors reviewed the results and approved the final version of the following manuscripts.

Introduction

Evidence-based practice (EBP) in healthcare is imperative for optimum patient care.[1] As part of the EBP process, clinicians must collect, appraise, and interpret clinical research to guide their clinical practice.[2] Evidence-based practice also encompasses clinical expertise (i.e., the proficiency and judgment acquired through experience and practice) and patient values.[2,3] As such, patient outcome measures may be collected and assessed to support EBP in the clinical setting[4] in the form of clinician-derived measures (e.g., strength assessment, range of motion measurements) or patient-reported outcome (PRO) measures which may be region-, patient-, disease-, or dimension-specific.[5-7] These PRO measures may be unidimensional (i.e., measuring physical *or* psychological constructs) or multidimensional (i.e., measuring physical *and* psychological constructs) instruments[8-10] used to capture the injury and recovery process from the patient's point of view.[1,4,11-13] Multidimensional PROs can be generic instruments (i.e., not specific to body location or injury) or region-specific scales (e.g., specific to the hip).[9,14] However, multidimensional region-specific instruments may also include assessment of health-related Quality of Life (QOL), which provides clinicians with valuable information on the patient's perception of the recovery process.[7,15]

Two PROs commonly used in patient care are the Hip Disability and Osteoarthritis Outcome Score (HOOS) and the Hip Disability and Osteoarthritis Outcome Score Joint Replacement (HOOS-JR). The HOOS is a 40-item instrument that was developed as a multidimensional region- and disease-specific outcome to assess hip disability pertaining to osteoarthritis (OA) and inury,[11] and is frequently used to evaluate outcomes after a total hip arthroplasty (THA).[11,13] In addition, the HOOS-JR was developed as a 6-item short-form version of the HOOS intended to be shorter and less burdensome compared to the 40-item HOOS, yet specific for patients undergoing a THA.[16] Both of these PRO measures are essential to clinicians to ensure the effectiveness of the surgical intervention used across the world.

Preliminarily analyses focused on the reliability, criterion validity and responsiveness of the HOOS and HOOS-JR.[11,16,17] However, despite the use of these outcomes in clinical practice, complete and robust psychometric analyses have minimally been assessed.[18] Psychometric testing, including confirmatory factor analysis (CFA), invariance testing, and latent growth-curve modeling are warranted to ensure the items within the instrument are appropriately measuring the constructs they intend.[19-21] When exploring the previously proposed HOOS and HOOS-JR, performing a CFA on the scale structure would ensure the items are consistent and measuring the uniqueness of the dimensions (e.g., questions pertaining to pain, QOL).[20,22] If the scale structure does not hold, modifications to the survey may be warranted. When modifying a survey, items should be evaluated for content, applicability and readability in the intended patient population.[23] The scale then must undergo exploratory factor analysis (EFA) to determine the number of dimensions that should be included.[20,24] In addition to the EFA procedures, inter-item correlations, correlations between constructs, and reliability (i.e., Cronbach's alpha and MacDonald's Omega) are assessed on the subscales of the alternate solution to determine if further modifications are warranted.[24] Once a parsimonious alternate solution is identified, CFA procedures are then assessed on the new scale to confirm the scale structure.[20,22]

Following CFA procedures, the alternate solution is subjected to invariance testing between patient groups (e.g., males and females) and across time (e.g., preoperative, postoperative).[19,20] An invariant model across subgroups ensures individuals from different groups are interpreting the survey items and meanings of the items similarly, regardless of the group classification.[19,22] This confirms scores from the instrument truly correspond with the construct and are not due to group specific designations.[19,22] In addition, longitudinal measurement invariance is important to determine if individual responses at each time point are representing a similar underlying construct (i.e., dimension), which is necessary to assess changes over time accurately.[19,22] Therefore, multi-group and longitudinal invariance is necessary to ensure the instrument can be used to compare hypothesized group differences to establish scale validity over time (i.e., patient visits). Lastly, once the scale is identified to be invariant, latent growth-curve (LGC) modeling can be used to understand how patients perceive their change in pain and level of function throughout the healing process. This modeling technique can be used to study the change in data when the outcome variable (e.g., HOOS-JR) is collected at multiple time points. In addition, the use of LGC allows for examination of intraindividual (within-person) change over time, as well as interindividual (between-person) variability.[19,25]

In a preliminary study, we assessed the scale structure of the HOOS and HOOS-JR in a mostly healthy population, however we failed to assess a large population of patients who underwent a THA. Our results indicated the CFA did not meet the recommended goodness-of-fit indices for the five-factor, 40-item HOOS, however, the HOOS-JR met most recommended fit indices. In addition, literature assessing longitudinal invariance and LGC modeling of the HOOS and HOOS-JR is lacking. Therefore, the purpose of this research was to assess the psychometric properties of the HOOS and HOOS-JR in a large diverse sample of patients who underwent a THA to determine its use in clinical practice and research.

References

1. Snyder AR, Parsons JT, Valovich McLeod TC, Curtis Bay R, Michener LA, Sauers EL. Using disablement models and clinical outcomes assessment to enable evidence-based athletic training practice, part I: disablement models. *J Athl Training.* 2008;43(4):428-436.

2. Hurley WL, Denegar CR, Hertel J. *Research methods: a framework for evidence-based clinical practice.* Wolters Kluwer/Lippincott Williams & Wilkins Health; 2011.

3. Sackett DL, Rosenberg WM, Gray J, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *Brit Med J.* 1996;312(7023):71.

4. Baker RT, Burton D, Pickering MA, Start A. Confirmatory Factor Analysis of the Disablement in the Physically Active Scale and Preliminary Testing of Short-Form Versions: A Calibration and Validation Study. *J Athl Train.* 2019;54(3):302-318.

5. Abrams D, Davidson M, Harrick J, Harcourt P, Zylinski M, Clancy J. Monitoring the change: current trends in outcome measure usage in physiotherapy. *Manual Ther.* 2006;11(1):46-53.

6. Valovich McLeod TC, Snyder AR, Parsons JT, Curtis Bay R, Michener LA, Sauers EL. Using disablement models and clinical outcomes assessment to enable evidence-based athletic training practice, part II: clinical outcomes assessment. *J Athl Train.* 2008;43(4):437-445.

7. Livermore-Brasher JS, Baker R, Nasypany A, Cheatham SW. Guidelines for Useful Integration of Patient-Rated Outcome Measures into Clinical Practice. *Clinical Practice in Athletic Training.* 2018;1(2):14-31.

8. Dobrzykowski E. The methodology of outcomes measurement. *Journal of Rehabilitation Outcomes Measurement.* 1997;1(1):8-17.

9. Jette AM. Outcomes research: shifting the dominant research paradigm in physical therapy. *Phys Ther.* 1995;75(11):965-970.

10. Watts JH, Clement DG. Conceptual framework for rehabilitation outcomes research. *Journal of Rehabilitation Outcomes Measurement.* 2000;4(2):55-61.

11. Nilsdotter AK, Lohmander LS, Klässbo M, Roos EM. Hip disability and osteoarthritis outcome score (HOOS)–validity and responsiveness in total hip replacement. *BMC Musculoskel Disord.* 2003;4(1):10.

12. Ware Jr JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36): I. Conceptual framework and item selection. *Med Care.* 1992:473-483.

13. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. 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.* 1988;15(12):1833-1840.

14. Vela LI, Denegar CR. The disablement in the physically active scale, part II: The psychometric properties of an outcomes scale for musculoskeletal injuries. *J Athl Train.* 2010;45(6):630.

15. Valier ARS, Jennings AL, Parsons JT, Vela LI. Benefits of and barriers to using patient-rated outcome measures in athletic training. *J Athl Train.* 2014;49(5):674-683.

16. Lyman S, Lee Y-Y, Franklin PD, Li W, Mayman DJ, Padgett DE. Validation of the HOOS, JR: a short-form hip replacement survey. *Clin Orthop Relat Res.* 2016;474(6):1472-1482.

17. Iorio R. Validation of the HOOS, JR: a Short-form hip replacement Survey. *Clin Orthop Relat Res.* 2016;474(6):1483-1485.

18. Miley EN, Casanova MP, Cheatham SW, Larkins L, Pickering MA, Baker RT. Confirmatory Factor Analysis of the Hip Disability and Osteoarthritis Outcome Score (HOOS) and Associated Sub-scales. *Int J Sports Phys Ther.* 2023;18(1):145.

19. Byrne BM. Structural equation modeling with AMOS: basic concepts, applications, and programming (multivariate applications series). *Taylor & Francis Group;* 2010.

20. Brown TA. *Confirmatory factor analysis for applied research.* Guilford publications; 2015.

21. Bryant FB, Yarnold PR. *Principal-components analysis and exploratory and confirmatory factor analysis*. American Psychological Association; 1995.

22. Kline RB. *Principles and practice of structural equation modeling.* Guilford publications; 2015.

23. Dillman DA, Smyth JD, Christian LM. *Internet, phone, mail, and mixed-mode surveys: the tailored design method.* John Wiley & Sons; 2014.

24. Leech NL, Barrett KC, Morgan GA. *IBM SPSS for intermediate statistics: Use and interpretation.* Routledge; 2014.

25. Preacher KJ, Wichman AL, MacCallum RC, Briggs NE. *Latent growth curve modeling.* Sage; 2008.

Chapter 1

Manuscript 1: Psychometric Analysis of the Hip Disability and Osteoarthritis Outcome Score Joint Replacement (HOOS-JR)

Abstract

Background: Within orthopedics, total hip arthroplasty (THA) is one of the fastest growing surgical procedures performed each year. Centers for Medicare and Medicaid Services (CMS) has requested hospitals collect and report patient-reported outcomes (PROs) beginning in 2024. More specifically in patients undergoing THAs, CMS has requested the collection of the Hip Disability and Osteoarthritis Outcome Score Joint Replacement (HOOS-JR). Therefore, the primary purpose of this study was to assess model fit of the HOOS-JR in a large, diverse sample of patients who underwent a THA to examine its psychometric properties. The secondary purpose was to conduct multi-group invariance testing of the scale between age groups and sex on the parsimonious scale structure identified.

Methods: The Surgical Outcomes System was retrospectively queried for the years 2014-2020 to establish a large data set of patients who underwent a THA to assess psychometric properties of the HOOS-JR. A confirmatory factor analysis (CFA) was performed on the 6-item HOOS-JR. Multi-group invariance testing was also performed on the sample of patients across sex and age groups.

Results: The one-factor, 6-item CFA did not meet the recommended fit indices for CFI (0.938), TLI (0.896), IFI (0.938), or RMSEA (0.143) criteria. A one-factor, 5-item CFA had acceptable fit for the sample data: all factor loadings were significant ($p \leq .001$) and goodness-of-fit indices met recommended values for CFI (0.981), TLI (0.961), IFI (.981), and SRMR (0.024); however, RMSEA (0.091) was slightly elevated. Invariance testing criteria were met between the age groups; scalar invariance was not met for sex.

Conclusion: The 6-item HOOS-JR did not meet contemporary model fit indices indicating that scale refinement is warranted. The 5-item met most goodness-of-fit indices and invariance criteria; as such, the 5-item HOOS-JR may be used to assess for differences between the different groups studied. However, further scale refinement may be warranted as localized fit issues, such as high error correlations, were identified. Scale modifications may be pertinent to develop a short-form instrument designed for multiple groups of patients undergoing a THA to allow clinicians and researchers the ability to confidently interpret these clinical outcomes.

Keywords: total hip arthroplasty, confirmatory factor analysis, invariance testing

Introduction

Within orthopedics, total hip arthroplasty (THA) is one of the fastest growing surgical procedures performed each year.[1,2] Between 2000 and 2014, there was an estimated 132% growth in the number of primary THAs performed in the U.S and THAs are projected to increase by 71.2% by 2030.[3] In addition, revision THAs are expected to grow between 43-70% by 2030.[2] The fastest growing groups to undergo this procedure include females and those aged 45-84 years.[4]

As the number of hip arthroplasties continue to rise, the Centers for Medicare and Medicaid Services (CMS) have requested hospitals collect patient-reported outcomes (PROs) beginning in 2024.[5] Specific to patients undergoing THA, CMS has requested the collection of the Hip Disability and Osteoarthritis Outcome Score (HOOS) Joint Replacement (JR).[5] Unlike many of the PROs that currently exist to measure hip disability and pain, the HOOS-JR survey was developed as an alternative PRO to be shorter and less burdensome compared to the 40-item HOOS, yet specific for patients undergoing a THA.[6]

The HOOS-JR was derived from the original HOOS, which consisted of 40 items in five domains: symptoms (S: five items); pain (P: 10 items); activity limitations-daily living (A: 17 items); sport and recreation (SP: four items); and hip-related QOL (Q: four items).[6-8] The development of the HOOS-JR began with removal of the QOL domain (i.e., four items) because it did not address activities or movements specific to the hip.[6] Mean relevant scores were then calculated for each remaining HOOS item; items with a score of 66.6% or less were removed.[6] Redundant items (i.e., going up or down stairs, walking on a flat surface, standing upright, and walking on an uneven surface) were also removed from the item pool.[6] A principle component factor analysis (PCA) was then conducted on the remaining 30 items, and a 30-item unidimensional solution was identified.[6] A Rasch analysis, which provides psychometric information about an instrument to facilitate logical and substantiated modifications, was performed on those items.[9] A final solution (i.e., the HOOS-JR) was accepted using only six items from the original 40-item: two items from the pain subscale and four items from the activity limitations-daily living subscale.[9]

Minimal psychometric examination of the HOOS-JR has been conducted since its creation, with most of the work focused on the validity, reliability, and responsiveness of the instrument. The HOOS-JR has been reported to have acceptable internal consistency (0.86-0.87) and high responsiveness (0.80).[6] Criterion validity, assessed using Spearman’s Correlation Coefficient, has been reported to be acceptable with moderate to high correlations with the HOOS subscales (0.60-0.94).[6,10] The HOOS-JR has also been considered reliable based on internal consistency values

(Cronbach's alpha = 0.86)[10]; values between 0.70 and 0.89 have been recommended to establish sound consistency without item redundancy.[11,12]

Although this initial work on the HOOS-JR provides some evidence for instrument reliability and validity, factorial validity and measurement invariance across subgroups have not sufficiently been established in the literature. Researchers have recently indicated the HOOS-JR met some, but not all, model fit recommendations (CFI = 0.965; TLI = 0.941; IFI = 0.965; RMSEA = 0.133) in a mostly healthy population who completed all 40 items of the HOOS.[13] Multi-group invariance testing on a small sub-sample of participants in this study who were diagnosed with hip osteoarthritis (OA) and/or previous history of a total hip replacement indicated the HOOS-JR may accurately capture valid group differences between those with hip injury and disability, and those who are healthy which provides initial support for scale validity.[13] However, the small sample sizes, utilization of healthy participants, and potential item response influence of including all 40 original HOOS items necessitates further psychometric testing of the HOOS-JR.

Thus, there is a need to complete more robust psychometric analysis of the HOOS-JR in a larger, more diverse sample of patients who have undergone THA. Further analysis steps include conducting a confirmatory factor analysis (CFA) to test the hypothesized factor structure of the HOOS-JR in an appropriate patient population and to conduct multi-group invariance tests to determine if the scale is generalizable and unbiased towards different groups. An invariant model across subgroups ensures individuals from different groups are interpreting the survey items and meanings of the items similarly, regardless of the group classification (e.g., male or female). This confirms scores from the instrument truly correspond with the construct and are not due to group specific designations; instrument multi-group invariance is necessary to ensure the instrument can be used to compare hypothesized group differences.[14-16] Therefore, the primary purpose of this study was to assess model fit of the HOOS-JR in a large, diverse sample of patients who underwent a THA to examine its psychometric properties. The secondary purpose was to conduct multi-group invariance testing of the scale between age groups and sex on the parsimonious scale structure identified.

Methods

Surgical Outcomes System (SOS; Arthrex, Inc., Naples, FL) was retrospectively queried for the years 2014-2020 to establish a large data set to assess the psychometric properties of the HOOS-JR. The SOS is a securely maintained database that contains information submitted from surgical centers worldwide. The University Institutional Review Board (IRB) indicated approval for this study was not required because analysis of the deidentified data set from the SOS database was not considered human subject research. However, IRB approval was granted by the Cedar-Sinai Office of Research Compliance and Quality Improvement as part of a larger research project using SOS data. All patients who underwent a THA and completed the HOOS-JR were queried for responses. De-identified responses to the HOOS-JR and necessary demographic information (e.g., sex, ethnicity, race, age at treatment) utilized for multi-group invariance testing were downloaded from the SOS for analysis in this study.

Hip Disability and Osteoarthritis Outcome Scale Joint Replacement

Participants rated how frequently they engaged in the behaviors over the past week using a 5-point Likert scale (0 = none, 1 = mild, 2 = moderate, 3 = severe, and 4 = extreme) for all six items.[6] The responses to each item (i.e., raw score) are first summed together on a range from 0 to 24, with 0 indicating perfect patient-perceived hip health.[6] Raw scores are then converted to an interval score ranging from 0 (raw score of 24) to 100 (raw score of 0).[6] Raw item scores were used for the purpose of this study.

Data Analysis

Data were exported from the SOS and downloaded using Microsoft® Excel for Mac (Version 16.46; Redmon, WA). Data were uploaded to Statistical Package for Social Sciences Version 28.0 (IBM Corp., Armonk, NY) for data cleaning and analysis. Missing data pertaining to the HOOS-JR were subject to the multiple imputation method; this process replaces the missing values with a set of random values based on the distribution of the sample.[17,18] Because the primary purpose was to assess the HOOS-JR, individuals with missing demographic data were not excluded from analysis and were left as missing values. Histograms, skewness, and kurtosis values were used to assess the normality of the data. Univariate outliers were assessed to determine if the z-scores exceeded the cut-off value of |3.3|. Multivariate outliers were also assessed using descriptive statistics and Mahalanobis distance of ≥ 16.81[15]; these data were identified using a chi-square table with degrees of freedom and *p*-value of 0.01.[15,19] This methodology generated the final data set used for the analysis.

Internal Consistency

Internal consistency was assessed by calculating Cronbach's alpha (α) and McDonald's maximum likelihood omega (ω) for the 6- and 5-item unidimensional scale.[12,14,15,20] An acceptable range for Cronbach's alpha and McDonald's maximum likelihood ratio was set at ≥ 0.70 and ≤ 0.89.[11, 20] Values < 0.70 indicate inadequate internal consistency, while values > 0.89 indicate item redundancy.[12,14,15,20]

Scale Structure - Confirmatory Factor Analysis

The final data set was used to conduct a CFA using Analysis of Moment Structures (AMOS) software (IBM Corp., Armonk, NY) on the 6-item HOOS-JR. As the HOOS-JR is scored as a one-factor model, the scale was defined as a one-factor, 6-item model.[6] Full Information Maximum Likelihood Estimation was used to generate the parameter estimates.[15,16] Model fit statistics included the likelihood ratio statistic (CMIN), Goodness of Fit Index (GFI), Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Bollen's Incremental Fit Index (IFI), Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Square Residual (SRMR).[15,16,21] Model fit was evaluated based on *a* priori values: GFI ≥ 0.95, CFI ≥ 0.95, TLI ≥ 0.95, RMSEA ≤ 0.06, IFI; ≥ 0.95, SRMR ≤ 0.08.[15,16,21-25] Given the influence small *df* can have on the RMSEA (i.e., potential model misfit),[21-24] the SRMR and CFI held greater weight in decisions regarding model fit testing. While the RMSEA value was elevated in both models, the value for the SRMR (i.e., 6-item = 0.043; 5-item = 0.024) met the recommended criterion (≤ 0.08).[24,25] In addition to assessing the overall goodness of fit, the interpretability, size, and significance of the model's parameter estimates (i.e., factor variances, covariances, and indicator errors) were reviewed to identify any localized areas of strain.[14]

Multi-group Invariance Testing

The full sample was then subjected to multi-group invariance testing where appropriate based on CFA results.[14,15,26] AMOS software was utilized to perform the analysis across sex (i.e., male, female) and age (i.e., ≤ 44, 45 to 54, 55 to 64; 65 to 74, ≥ 75) subgroups.[14,26,27] This was accomplished using a set of hierarchical procedures with an increasing level of constraint to determine if the respective items of the construct were stable and approximately equal across groups.[14,15,26]

The model then underwent configural, metric, and scalar invariance testing.[14,15,26] First, the configural invariance test placed all groups in the same model to ensure the same factors have similar items across sub-groups. Secondly, the metric model then tested if factor loadings were equal across sub-groups.[14,16] If the model met metric invariance requirements, equal variances (i.e., group differences) between groups were then assessed.[14,16] Lastly, the scalar invariance test ensured that item intercepts were equal across groups, which indicated the means were not determined or altered by external factors.[14,16] If the model met scalar invariance requirements, equal mean models (i.e., score differences) were tested between groups.[16]

Model fit was compared using the CFI difference test (CFI_{DIFF}) and the chi-square difference test (χ^2_{DIFF}), with a *p*-value cut-off of 0.01.[22,26] Given the sensitivity of the χ^2_{DIFF} test to sample size,[22] the CFI_{DIFF} test held greater weight in decisions regarding invariance testing model fit. If a model exceeded the χ^2_{DIFF} test, but met the CFI_{DIFF} test, invariance testing continued.

Results

A total of 12640 patient responses were exported from the SOS database and were included in the analysis. A total of 425 patients were identified as multivariate outliers (i.e., Mahalanobis ≥ 16.81) and were removed from the analysis. Of the patients identified as outliers, 240 identified as females and 136 identified as males. Those deleted represented each age group (i.e., ≤ 44, 45-54, 55-64, 65-74, ≥ 75) with a mean age of 65.89 ± 12.29 (standard deviation; SD). Upon examination of the skewness and kurtosis of the sample, no items had a non-normal distribution (i.e., < 3.3). The final sample comprised of 12215 patients with an age range from 16 to 90 years (mean age = 63.39 ± 10.84 years; median age = 64.0 years), with females accounting for 49.5% (n = 6051) of the sample (Table 1). In the sample, the largest age group represented was 65-74 (31.7%) followed by the 55-64 group (31.2%) (Table 1).

Internal Consistency

Cronbach's alpha (α) and McDonald's maximum likelihood omega (ω) was conducted on the full sample of patients undergoing a THA at baseline (i.e., preoperative). Cronbach's alpha (α) and McDonald's omega (ω) were acceptable for the 6-item unidimensional scale (α = 0.88; ω = 0.88). In addition, Cronbach's alpha (α) and McDonald's omega (ω) were acceptable for the 5-item unidimensional scale (α = 0.86; ω = 0.86).

Scale Structure - Confirmatory Factor Analysis

The one-factor, 6-item CFA of the HOOS-JR did not meet the recommended fit indices for the CFI (0.938), TLI (0.896), IFI (0.938), or RMSEA (0.143) (Figure 1); SRMR did meet recommended levels (0.043). Factor loadings ranged from 0.71 to 0.78 and were significant ($p \leq$ 0.001). Modification indices revealed significant and meaningful cross-loadings between several items were present. Additionally, high error-term correlations were identified between items one and two (i.e., item one [going up or down stairs], item two [walking on an uneven surface];1476.92); also, between items five and six (i.e., item five [lying in bed turning over, maintaining hip position], item six [Sitting]; 584.39).

As there was a high error-term correlation identified between items one and two, in addition to the lack of overall-fit, we explored the removal of item one. The one-factor 5-item CFA of the HOOS-JR met the recommended fit indices for the CFI (0.981), the TLI (0.961), and the IFI (0.981) (Figure 2). The RSMEA value did not meet the recommended model fit guidelines (0.091); however, SRMR met the recommended model fit indices recommendations (0.024). Modification indices revealed significant and meaningful cross-loadings between several items were present. Additionally, there was a high error-term correlation present between items five and six (i.e., item five [lying in bed turning over, maintaining hip position], item six [Sitting]; 328.32)

Multi-group Invariance Testing

Sex Subgroups

Invariance testing was performed on the sample of patients who either identified as male or females (n = 11479), with missing data excluded from the analysis. The initial model (configural) demonstrated acceptable model fit (CFI = 0.980; χ^2 [4] = 499.93; RMSEA = 0.065; Table 2), indicating equal form across sex. The metric model (i.e., equal loadings) passed both the CFI_{DIFF} test and the χ^2_{DIFF} test; as the model satisfied invariance criteria, the equal latent variance model was conducted. Both the CFI_{DIFF} test and the χ^2_{DIFF} test both met invariance, indicating variances were equal across sex. The scalar model (i.e., equal loadings and intercepts) only slightly exceeded the CFI_{DIFF} test (i.e., 0.013); as such, we explored the latent means across sex. The equal latent means model did not meet model fit criteria; further assessment of the means indicated females reported significantly higher scores (i.e., more hip disability) than males.

Age Subgroups

Invariance testing was performed on the sample of patients with a reported age (n = 11725); those with a missing age were excluded from the analysis. The initial model (configural) demonstrated acceptable model fit (CFI = 0.980; χ^2 [20] = 522.74; RMSEA = 0.041; Table 3), indicating equal form across age groups. The metric model (i.e., equal loadings) passed both the CFI_{DIFF} test and the χ^2_{DIFF} test; as the model satisfied invariance criteria, the equal latent variance model was conducted. Both the CFI_{DIFF} test and the χ^2_{DIFF} test both met invariance criteria, indicating variances were equal across age groups. The scalar model (i.e., equal loadings and intercepts) passed both the CFI_{DIFF} test and the χ^2_{DIFF} tests; as such, the model allowed for comparison of the latent means across age groups. The equal latent means model did not meet model fit criteria; further assessment of the means indicated those who were in the age group of 55-64, 65-74, and > 75 reported significantly lower scores (i.e., less hip disability) than those in the age groups < 45 and 45-54.

Discussion

The purpose of this study was to assess the scale structure and multi-group invariance testing (i.e., sex and age groups) using CFA procedures of the 6-item HOOS-JR, as this outcome is commonly used in in clinical practice following THA. Contemporary CFA procedures provide a more rigorous examination of the HOOS-JR for model fit and multi-group invariance. Overall, the 6-item model did not meet fit recommendations; however, the 5-item model met most contemporary fit recommendations. It may be prudent to alter the original 6-item HOOS-JR to ensure that a suitable tool exists to measure pain and functional disability following a THA.

Scale Structure - Confirmatory Factor Analysis

When defined as a one-factor, 6-item structure, the model did not meet the recommended overall goodness-of-fit (i.e., model fit indices) at baseline (i.e., preoperative). Model fit was poor (CFI = 0.938; TLI = 0.896; IFI = 0.938; RMSEA = 0.143; SRMR = 0.043), with specific concerns related to item correlations and error-term covariances. The findings in this study are different than previous findings[13]; previous model fit criteria met some model fit indices (CFI = 0.965; TLA = 0.941; IFI = 0.965).[13] However, the previous study included a population sample of mostly healthy and younger individuals, whereas the present study includes a population sample whose mean age is higher and those undergoing a THA.

Modification indices revealed significant covariances between error terms (i.e., items one and two; items five and six). Moreover, upon further review of the items, item one (i.e., going up or down stairs) and item two (i.e., walking on an unstable surface) had a significantly high error-term correlation (1476.92) indicating overlap between the two items[15,16]; this finding is similar to previous literature.[13] Additionally, item one is considered double barreled, meaning that the item may be asking the patient to report about their pain performing more than one activity (i.e., going up or down stairs).[28]

Because item one was identified to have a high meaningful error-term covariance in multiple studies, and the item structure is considered double-barreled,[28] we determined there was theoretical support for removal of this item. Upon removal, the defined one-factor, 5-item structure met most recommended model fit indices (CFI = 0.981; TLI = 0.961; IFI = 0.981). Upon assessment of the modification indices, the high error-term covariance (328.32) was still identified between item five (i.e., lying in bed [turning over, maintaining hip position]) and item six (sitting).

While this analysis met most of the goodness-of-fit indices, the RMSEA was higher than the recommended "best practices" value (i.e., ≤ 0.06) in the 5-item (i.e., 0.091) model[14,15,24,25]; however, it has been recommended that RMSEA be interpreted with caution with small *df* models due to the potential for incorrectly rejecting a correctly specified model.[21,24] Recently, it has been recommended in the literature to report the SRMR instead of RMSEA in models with small *df*.[21,24] As such, we decided to report both the RMSEA and the SRMR given the small *df* in our model. While the RMSEA value was elevated in both models, the value for the SRMR (i.e., 6-item = 0.04; 5-item = 0.02) met the recommended criterion (≤ 0.08).[24,25]

Findings from these analyses indicate that modifications to the scale are warranted. As the HOOS-JR is required by CMS to be collected in patients undergoing a THA,[29] a psychometrically sound instrument is needed for clinicians to accurately draw conclusions on outcomes within their clinical practice, and for CMS to have an instrument for accurate reimbursement.[29,30] Therefore, modifications to the HOOS-JR, such as rewording items, removing items, or creating new items, are necessary to create a PRO that effectively measures hip disability and pain.[28]

As the HOOS-JR 6-item model did not meet recommended model fit criteria, further testing (i.e., multi-group invariance testing) is not recommended.[14-16] Additionally, it remains prudent to consider altering the instrument prior to its continued use to ensure clinicians are capturing appropriate surgical outcomes in a variety of patient populations. However, due to the frequent and necessary use of collecting the HOOS-JR in clinical practice for clinicians currently, we determined that there is a need to better understand the scale. As the 5-item model met most model fit indices, multi-group invariance testing of the one-factor 5-item model was performed.

Multi-group Invariance Testing

Invariance testing may be conducted for several reasons including, but not limited to, the assessment of an instrument's items to ensure similar interpretation across groups (e.g., age groups, sex), assessment of the underlying construct (e.g., hip disability) to ensure similar measurement across groups, or assessment of measurement properties across repeated measures (i.e., longitudinal).[15,16] As such, multi-group invariance testing was conducted to determine if the HOOS-JR items were being interpreted similarly across subgroups (i.e., sex and age) and if the construct (i.e., hip disability) is being measured similarly across groups.[15,16] To our knowledge, this was the first study to perform multi-invariance procedures on the HOOS-JR in a pure sample of joint replacement patients. Additionally, given its heavy use in clinical practice, conducting invariance testing across multiple groups is pertinent to determine if the scale is adequately measuring patient's perceived hip disability across groups.

The initial (i.e., CFA) results indicated that the one-factor, 5-item model met the stringent goodness-of-fit standards[14,15]; however, understanding the outcome in its current form (i.e., 6-item scale) is essential while further scale modifications occur. Concerns with the 5-item HOOS-JR were identified when multi-group invariance testing between sex was performed. Baseline models for males and females were performed; males met some model fit indices (CFI = 0.97, TLI = 0.95; IFI = 0.97; RMSEA = 0.11; SRMR = 0.03), females met model fit indices (CFI = 0.99; TLI = 0.97; IFI = 0.99; RMSEA = 0.08; SRMR = 0.02). The configural model (i.e., equal form) meet model fit recommendations, meaning group differences in variances were not noted in pain and function between sex. As the scalar model only slightly exceeded contemporary model fit recommendations, equal means model was explored. When the means were constrained to be equal, the model exceeded the recommended model fit indices; further assessment revealed that females report more pain and less function when compared to males. Due to the lack of scalar invariance, these findings indicate that males and females do not conceptualize hip disability similarly when using the 5-item HOOS-JR.

Multi-group invariance testing was also performed on the 5-item HOOS-JR across several age groups. Our results identified that the 5-item HOOS-JR is invariant across the examined age groups which indicated the scale can be used to assess age group differences in hip pain and function across these groups. Additionally, we found statistically significant latent mean differences between two specific age groups, with those aged 55-90 reported lower scores (i.e., less hip disability) than those aged 19-54. This is consistent with other authors who report two distinct patterns of activity levels (i.e., ≥ 55 and < 55) pre- to postoperative, with those < 55 reporting lower activity levels postoperatively when compared to the older group.[31] Additionally, several factors (e.g., level of activity, number of comorbidities, body mass index, American Society of Anesthesiologists score, medication use, etc.) have been identified in older populations (i.e., older than 55) that could explain the findings.[31,32] Specifically, these factors may contribute to older patients being less active which may then result in lower self-reported scores.

Even though our data represents a large sample of patients who underwent a THA, we were unable to obtain the associated diagnoses in this population. Therefore, it should be recognized that regardless of the diagnoses (e.g., OA, osteonecrosis, periprosthetic joint infection), comorbidities, medication use, etc., patients interpreted the questions similarly across the age groups and the means are not driven or contaminated by outside factors (e.g., cultural norms, group specific attributes).

However, further research is needed to understand these age group differences on the outcomes pre- and post-surgical intervention.

Limitations and Future Research

Several limitations are present with our study. While a diverse population was assessed globally, most of the responses were collected from the United States (i.e., 12197). Thus, future research should include psychometric assessment of the scale in samples across other countries as well to ensure appropriate measurement properties across cultures and languages. Additionally, upon export of the sample, we were unable to capture race and ethnicity due to a high amount of missing data (i.e., greater than 10000 cases). As such, we were unable to assess the psychometrics of the scale across these groups within the sample. Future research should conduct invariance testing across race and ethnicity to ensure the scale has the necessary properties to support between groups analysis in these populations. Further, we were unable to determine if the HOOS-JR was collected across different medical history, and surgical (e.g., laterality, approach, primary vs. revision, inpatient vs. outpatient) or diagnostic (e.g., grade of OA, periprosthetic joint infection) variables due to the limitations of the database; thus, caution is warranted when examining HOOS-JR differences in groups that have not been analyzed. As this instrument is typically collected across time to assess outcomes following a THA, future research should also be conducted using similar methods to confirm the psychometric properties of the one-factor, 5-item HOOS-JR across time (e.g., longitudinal invariance testing) to ensure the scale can be used to assess change with repeated measures).

Conclusion

Structural properties of the 6-item HOOS-JR were assessed in patients at preoperative intervention and across groups (i.e., age and sex). Our findings do not support the use of the one-factor, 6-item HOOS-JR in clinical practice and research. However, structural and multi-group invariance properties of the one-factor, 5-item HOOS-JR meets most model fit indices and supports the use of the scale to assess differences between different age groups. Further scale modifications are warranted to develop a structurally sound PRO sensitive to measure change in in patients undergoing a THA. In its current form, the 6-item HOOS-JR should be used with caution in clinical practice and research.

References

1. Fingar KR, Steiner CA , Stocks C, Weiss AJ. Most Frequent Operating Room Procedures Performed in U.S. Hospitals, 2003–2012. In: Healthcare Cost and Utilization Project (HCUP) Statistical Briefs. Agency for Healthcare Research and Quality (US): Rockville, MD; 2006.

2. Bradbury Jr TL, Farley KX, Guild GN, Schwartz AM. Projections and epidemiology of revision hip and knee arthroplasty in the United States to 2030. *J Arthroplasty*. 2020;35(6):S79-S85.

3. Premkumar A, Sheth NP, Sloan M. Projected volume of primary total joint arthroplasty in the US, 2014 to 2030. *J Bone Joint Surg.* 2018;100(17):1455-60.

4. Chen L, Cleveland JD, Singh JA, Yu S. Rates of total joint replacement in the United States: future projections to 2020–2040 using the national inpatient sample. *J Rheumatol.* 2019;46(9):1134-40.

5. FY 2023 Hospital Inpatient Prospective Payment System (IPPS) and Long-Term Care Hospital Prospective Payment System (LTCH PPS) Final Rule — CMS-1771-F 2022. Available at: https://www.cms.gov/newsroom/fact-sheets/fy-2023-hospital-inpatient-prospective-payment-system-ipps-and-long-term-care-hospital-prospective. Accessed January 25, 2023.

6. Franklin PD, Lee Y-Y, Li W, Lyman S, Mayman DJ, Padgett DE. Validation of the HOOS, JR: a short-form hip replacement survey. *Clin Orthop Relat Res*. 2016;474(6):1472-82.

7. Bremander A, Nilsdotter 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. *Arthrit Care Res*. 2011;63(S11):S200-S7.

8. Klässbo M, Lohmander LS, Nilsdotter AK, Roos EM. Hip disability and osteoarthritis outcome score (HOOS)–validity and responsiveness in total hip replacement. *BMC Musculoskel Dis*. 2003;4(1):10.

9. Annetta L, Lamb RL, Vallett D. Development of a short-form measure of science and technology self-efficacy using Rasch analysis. *J Sci Educ Technol.* 2014;23(5):641-57.

10. Franklin PD, Gandek B, Roos EM, Ware Jr JE. A 12-item short form of the Hip disability and Osteoarthritis Outcome Score (HOOS-12): tests of reliability, validity and responsiveness. *Osteoarthritis Cartilage*. 2019;27(5):754-61.

11. Barrett KC, Leech NL, Morgan GA. *IBM SPSS for intermediate statistics: Use and interpretation*. Routledge; 2014.

12. Taber KS. The use of Cronbach's alpha when developing and reporting research instruments in Science Education. *Research in Science Education*. 2018;48(6):1273-96.

13. Miley EN, Casanova MA, Cheatham SW, Larkins L, Pickering MA, Baker RT. Confirmatory Factor Analysis of the Hip Disability and Osteoarthritis Outcome Score (HOOS) and Associated Sub-scales. *Int J Sports Phys Ther.* 2023.

14. Brown TA. *Confirmatory factor analysis for applied research*. Guilford publications; 2015.

15. Kline RB. *Principles and practice of structural equation modeling*. Guilford publications; 2015.

16. Byrne BM. *Structural equation modeling with AMOS: basic concepts, applications, and programming (multivariate applications series)*. Taylor & Francis Group; 2010.

17. Houck PR, Koru-Sengul T, Mazumdar S, Mulsant BH, Pollock BG, Tang G. Estimating treatment effects from longitudinal clinical trial data with missing values: comparative analyses using different methods. *Psychiatry Res.* 2004;129(2):209-15.

18. Olsen MK, Schafer JL. Multiple imputation for multivariate missing-data problems: A data analyst's perspective. *Multivariate Behav Res.* 1998;33(4):545-71.

19. Fidell LS, Tabachnick BG. *Using multivariate statistics.* Allyn & Bacon; 2001.

20. Streiner DL. Starting at the beginning: an introduction to coefficient alpha and internal consistency. *J Pers Assess.* 2003;80(1):99-103.

21. DiStefano C, Lee T, Maydeu-Olivares A, Shi D,. Evaluating SEM model fit with small degrees of freedom. *Multivariate Behav Res.* 2020:1-36.

22. Grimm LG, Yarnold PR. Reading and understanding multivariate statistics. American Psychological Association; 1995.

23. Coughlan J, Hooper D, Mullen M. Evaluating model fit: a synthesis of the structural equation modelling literature. 7th European Conference on Research Methodology for Business and Management Studies. 2008:195-200.

24. Taasoobshirazi G, Wang S. The performance of the SRMR, RMSEA, CFI, and TLI: An examination of sample size, path size, and degrees of freedom. *Journal of Applied Quantitative Methods.* 2016;11(3):31-9.

25. Lee T, Maydeu-Olivares A, Shi D. Understanding the model size effect on SEM fit indices. *Educ Psychol Meas.* 2019;79(2):310-34.

26. Byrne BM. *Structural equation modeling with Mplus: Basic concepts, applications, and programming*. Routledge; 2013.

27. Bercovitz A, Palso K, Wolford ML. *Hospitalization for total hip replacement among inpatients aged 45 and over: United States, 2000-1010.* US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics; 2015.

28. Christian LM, Dillman DA, Smyth JD. *Internet, phone, mail, and mixed-mode surveys: the tailored design method*. John Wiley & Sons; 2014.

29. Iorio R. Validation of the HOOS, JR: a Short-form hip replacement Survey. *Clin Orthop Relat Res.* 2016;474(6):1483-5.

30. Comprehensive Care for Joint Replacement Model. 2023 CMS. Available from: https://innovation.cms.gov/innovation-models/cjr. Accessed January 25, 2023.

31. Bonvin A, Lübbeke A, Roussos C, Perneger T, Stern R, Zimmermann-Sloutskis D. Physical activity before and after primary total hip arthroplasty: a registry-based study. *Arthritis Care Res.* 2014;66(2):277-84.

32. Amstutz HC, Higa JK, McClung CD, Zahiri CA. Schmalzried TP. Relationship between body mass index and activity in hip or knee arthroplasty patients. *J Orthop Res.* 2000;18(1):35-9.

Tables

Table 1

Demographic Information of Full Sample

	Full Sample (n = 12215)
Sex (%)	
Male	5428 (44.4)
Female	6051 (49.5)
Missing	736 (6.0)
Age ± standard deviation	63.39 ± 10.84
Age Group (%)	
44 and younger	553 (4.5)
45-54	1742 (14.3)
55-64	3813 (31.2)
65-74	3878 (31.7)
75 and older	1739 (14.2)
Missing	490 (4.0)
Ethnicity (%)	
Hispanic	85 (0.7)
Not Hispanic	1682 (13.8)
Patient declines to answer	38 (0.3)
Missing	10410 (85.2)

Table 2

Goodness-of-Fit Indices for Measurement Invariance of the 5-item HOOS-JR Across Sex

Modified 5-item HOOS-JR	χ^2	**df**	χ^2_{diff} **(df_{diff})**	**CFI**	**CFI_{diff}**	**TLI**	**RMSEA**
Males (n = 5428)	322.51	5	----	0.973	----	0.946	0.108
Females (n = 6051)	173.70	5	----	0.987	----	0.974	0.075
Configural (equal form)	496.21	10	----	0.980	----	0.960	0.065
Metric (equal loadings)	499.93	14	3.72 (4)	0.980	<0.001	0.972	0.055
Equal factor variances*	500.96	15	4.75 (5)	0.980	<0.001	0.980	0.053
Scalar (equal indicator intercepts)	831.34	18	**335.13 (8)**	0.967	**0.013**	0.963	0.067
Equal latent means*	919.65	19	**423.44 (9)**	0.963	**0.017**	0.961	0.064

* = Substantive questions; **Bolded** = did not meet cuff off criteria

Table 3

Goodness-of-Fit Indices for Measurement Invariance of the 5-item HOOS-JR Across Age

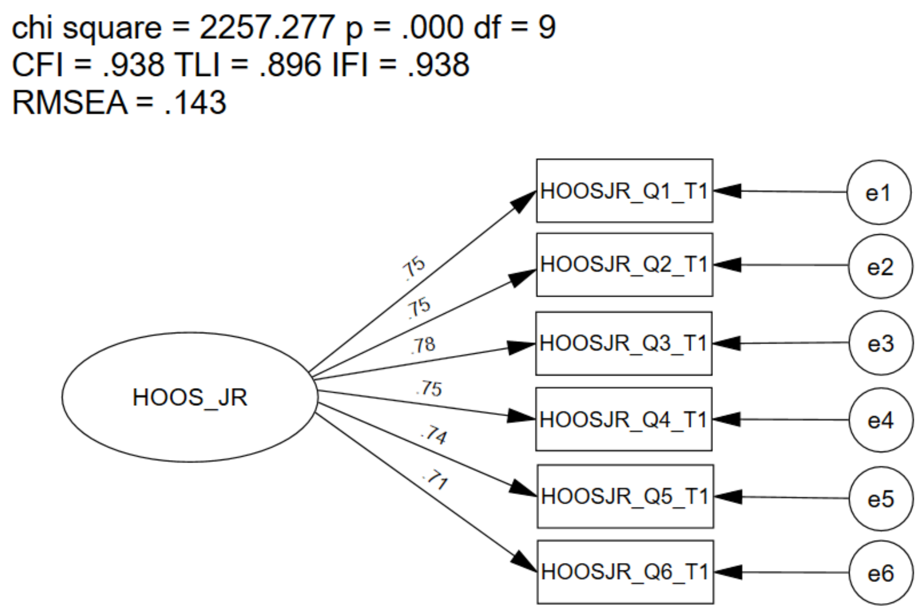
Modified 5-item HOOS-JR	χ^2	df	χ^2_{diff} (df_{diff})	CFI	CFI_{diff}	TLI	RMSEA
44 and younger (n = 553)	21.00	5	----	0.973	----	0.987	0.076
45-54 (n = 1742)	95.14	5	----	0.974	----	0.948	0.102
55-64 (n = 3813)	193.17	5	----	0.978	----	0.955	0.099
65-74 (n = 3878)	161.70	5	----	0.980	----	0.960	0.090
75 and older (n = 1739)	54.73	5	----	0.988	----	0.975	0.076
Configural (equal form)	525.74	25	----	0.980	----	0.960	0.041
Metric (equal loadings)	540.62	44	14.88 (19)	0.980	<0.001	0.976	0.032
Equal factor variances*	559.59	45	33.88 (20)	0.979	0.001	0.977	0.031
Scalar (equal indicator intercepts)	695.00	57	169.26 (32)	0.974	0.006	0.978	0.031
Equal latent means*	847.71	61	**321.97 (36)**	0.968	**0.012**	0.974	0.033

* = Substantive questions; **Bolded** = did not meet cuff off criteria

Figures

Figure 1

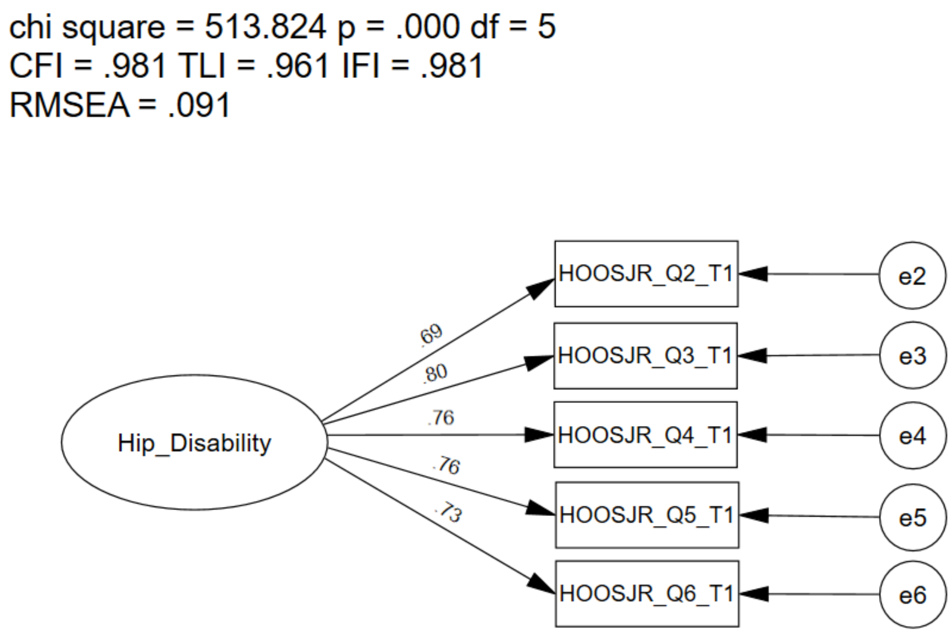
Confirmatory Factor Analysis of the One-Factor 6-item HOOS-JR



Chisq = Chi Square (χ^2); df = degrees of freedom, p = alpha level; CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; IFI = Bollen's Incremental Fit Index; RMSEA = Root Mean Square Error of Approximation; SRMR = Standardized Root Mean Square Residual.

Figure 2

Confirmatory Factor Analysis of the One-Factor, 5-item Alternate HOOS-JR Model



Chisq = Chi Square (χ^2); df = degrees of freedom, p = alpha level; CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; IFI = Bollen's Incremental Fit Index; RMSEA = Root Mean Square Error of Approximation; SRMR = Standardized Root Mean Square Residual.

Chapter 2

Manuscript 2: Longitudinal Analysis and Latent Growth Modeling of the Modified Hip Disability and Osteoarthritis Outcome Score Joint Replacement (HOOS-JR)

Abstract

Background: The Hip Disability and Osteoarthritis Outcome Score Joint Replacement (HOOS-JR) was developed as a short-form from the 40-item HOOS to measure progress after total hip arthroplasty (THA). However, longitudinal validity of the scale structure and latent growth curve (LGC) modeling (i.e., assessment of patients change in scores) pertaining to the modified 5-item HOOS-JR have not been assessed. Therefore, the purpose of the study was to evaluate the structural validity, longitudinal invariance properties, and LGC modeling of the modified 5-item HOOS-JR in a large sample of patients who underwent a THA.

Methods: A longitudinal study was conducted using the Surgical Outcome System (SOS) database. Longitudinal data from patients who underwent a THA were queried and included in our analyses. Confirmatory factor analyses (CFA) were conducted to assess the structural validity and longitudinal invariance across five time points (preoperative, 6-month-, 1-year-, 2-year- and 3-year postoperatively). In addition, LGC modeling was performed to assess the heterogeneity of the recovery patterns for different subgroups (i.e., age groups and sex) of patients. *A priori* cut-off values included the Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), and Bollen's Incremental Fit Index (IFI) ≥ 0.95, Root Mean Square Error of Approximation (RMSEA) ≤ 0.05, and Standardized Root Mean Square Residual (SRMR) ≤ 0.08.

Results: Results of the CFA met most of the goodness-of-fit indices at all time points. Longitudinal analysis did not meet full invariance; upon further assessment, item five was identified as problematic. Partial invariance requirements were met upon release of the intercept constraint associated with item five. The equal means model did not pass recommended goodness-of-fit indices. When means were not constrained to be equal, scores significantly changed over time ($p < 0.001$) with the highest scores on the modified 5-item HOOS-JR identified at preoperative and the lowest scores identified at 2- and 3-years postoperatively. In addition, mean scores significantly changed from preoperatively to 3-years postoperatively. Lastly, there was significant differences identified between sex groups over time.

Conclusion: Partial scalar invariance was identified by releasing the scores pertaining to item 5. In addition, we identified that patients self-report most improvement in their scores within the first 6-months postoperative. Females reported more hip disability at preoperative and had faster improvement measured by scores of the modified 5-item HOOS-JR. Significant differences were not found between the age groups.

Keywords: total hip arthroplasty, invariance testing, latent growth curve modeling

Introduction

Total hip arthroplasties (THAs) are considered a highly effective treatment for patients who have been diagnosed with joint deterioration disorders, such as osteoarthritis (OA), rheumatoid arthritis, or traumatic arthritis.[1-4] Nearly 300,000 THAs are performed annually in the United States, with rates predicted to grow by 174% from 2005 to 2030 due to increases in the population and disease diagnoses.[3,5,6] In addition, a large percentage (i.e., 70-85%) of THAs are performed on patients who have been diagnosed with arthritis which is often associated with ageing.[4,7] Not surprisingly, however, patients younger than 65 years of age have represented the fastest growing age group for THAs, accounting for nearly 47% of all THAs performed in 2012 compared to only 34% in 1997.[3,5,6] Current researchers have projected more than 50% of all THAs will be performed on patients younger than 65 years by 2030.[6,8,9] Thus, THAs are no longer only associated with the ageing population.[10]

Candidates for THAs often complain of pain and restricted range of motion (ROM) while performing activities[4,11]; as such, the goal of treatment is to reduce pain, improve ROM, and enhance overall quality of life (QOL).[2-4,10] Many THA patients have failed to experience substantial improvement in their condition from conservative treatment approaches, which often leads to the decision to consider a THA.[2-4,10] Surgical and implant technologies have continued to evolve producing excellent long-lasting outcomes that supports THA as a treatment option in the younger population.[7,10,12,13] Patient-reported outcomes (PROs) are often used before and after THA to measure pain, functionality, and QOL to assess surgical outcomes. The use of PROs helps clinicians and researchers determine the effectiveness of the treatment.[14] Due to the importance of outcomes to clinicians, researchers, and patients, it is imperative that PROs are valid, reliable, and responsive to change[14,15]; as such, PROs also need to be evaluated over a prolonged period of time.[16,17]

Outcomes currently used in the patient population with hip pain and dysfunction include the Harris Hip Score (HHS), Western Ontario & McMaster Universities Arthritis Index (WOMAC), Oxford Hip Score (OHS), the Hip disability and Osteoarthritis Outcome Score (HOOS), and more recently, the HOOS-JR (joint replacement) short-form.[18-25] The original one-factor, 6-item HOOS-JR was derived from the 40-item HOOS and developed as a short-form to measure outcomes specifically after surgical intervention; as such, the 6-item HOOS-JR is primarily used with pre- and postoperative THA.[26] Psychometric assessment of the HOOS-JR using contemporary methods has been minimally reported in the literature. Most clinometric assessment has focused on scale internal consistency, responsiveness of the instrument over three time points (i.e., preoperative, 6-months postoperative, and 12 months postoperative), and construct validity.[26,27] The 6-item HOOS-JR has been reported to have acceptable internal consistency (0.86-0.87),[26] and also high responsiveness (0.80).[26] Construct validity was established for the 6-item HOOS-JR by correlating it to the HOOS-pain subscale (0.87), ADL subscale (0.94), symptoms subscale (0.62), sport and recreation subscale (0.65), and the QOL subscale (0.60).[26,27] Also, the HOOS-JR was highly correlated to the WOMAC function (0.94), pain (0.84), and stiffness (0.64) subscales.[26] Internal consistency of the HOOS-JR has also been established (Cronbach's alpha = 0.86)[27]; values ranging 0.70-0.89 have been recommended to establish consistency without item redundancy.[28,29]

Further research pertaining to scale validity (i.e., factor validity) on the HOOS-JR is needed.[30] This process should be performed to ensure the scale is appropriate for clinical practice and research.[31,32] To date, a limited number of studies have been conducted to assess the psychometrics of the 6-itm HOOS-JR using contemporary procedures, such as confirmatory factor analysis (CFA), exploratory factor analysis (EFA) or principle component factor analysis (PCA) methods.[30-32] Lyman et al.[26] performed a PCA to determine the dimensionality of the items; authors identified that the 30 items, which were derived from the 40-item HOOS, were of a single construct. Further, this process in combination with a Rasch analysis, reduced the number of items (n = 40) of the HOOS to form the 6-item HOOS-JR.[26]

More recently, authors performed a CFA on the 6-item HOOS-JR identifying a promising factor structure which met most, but not all, of the model fit criteria (CFI = 0.965; TLI = 0.941; IFI = 0.965; RMSEA = 0.133) in a mostly healthy population.[33] Further, initial multi-group invariance testing was conducted; however, the analyses performed were only assessed at one time point and in a small sample (n = 656).[33] In a more recent study by E.N. Miley et al. (unpublished data, 2023), a CFA was performed on the 6-item HOOS-JR in a large sample of patients who underwent a THA (n = 12215). Goodness-of-fit indices were not met for the original 6-item HOOS-JR (CFI = 0.938; TLI = 0.896; IFI = 0.938; RMSEA = 0.143); further exploration identified a more parsimonious scale structure (CFI = 0.981; TLI = 0.961; IFI = 0.981; RMSEA = 0.091) including five items of the original HOOS-JR (i.e., items two-six). As such, the modified 5-item HOOS-JR warranted further investigation to confirm the psychometric properties pertaining to longitudinal use (i.e., outcome assessed over time).

Longitudinal research assessing the measurement properties of the modified 5-item HOOS-JR over the course of treatment and long-term follow-up is also lacking. Valid instruments to assess change over prolonged periods of time are needed postoperative THA because variations in outcomes may not emerge for years after the procedure.[16,17] Establishing measurement properties through invariance testing ensures that the interpretations across time are valid and reliable,[28,34] and longitudinal measurement invariance assessment is conducted to determine if individual responses at each time point are representing a similar underlying construct, which is necessary to assess changes over time accurately.[31,35] Thus, further measurement research on the modified 5-item HOOS-JR is needed to establish scale validity over time for its clinical use. The multi-step process should include conducting a CFA to examine the factor structure of the 5-item HOOS-JR at each time point, then conducting longitudinal measurement invariance testing (i.e., preoperatively, postoperatively).[31,34]

Further, assessing longitudinal data of the 5-item HOOS-JR may is important for clinicians to understand how patients perceive their change in pain and level of function during the healing process. Latent growth curve (LGC) modeling is a statistical technique designed to study the change in data when the outcome variable (e.g., 5-item HOOS-JR) is collected at multiple time points.[35-38] The use of LGC allows for examination of intraindividual (within-person) change over time, as well as interindividual (between-person) variability.[35,38] More recently, researchers have performed LGC to assess changes in PRO scores (i.e., OHS and HHS) pre- and postoperatively and noted that most of the healing occurs within the first 3- to 6-months postoperatively.[39-41] Not only is it imperative for clinicians to understand the rates at which patients heal, but also it is important to understand if differences exist in these rates of healing patterns based on age and sex. Studies assessing LCG modeling of any forms of the HOOS-JR or assessing if differences exist based on age and sex, are lacking in the literature.

Therefore, the purpose of the study was to evaluate the psychometric properties of the modified 5-item HOOS-JR in a three-step process: 1) a confirmatory factor analysis (CFA) of factor structure, using contemporary fit recommendations, in a large sample of patients who underwent a THA to ensure model fit at each time point (i.e., preoperatively, and 6-months, 1-year, 2-years, and 3-postoperatively); 2) longitudinal invariance testing (i.e., equal factor variances, equal factor covariance, and equal means) of the scale across multiple time points (i.e., i.e., preoperatively, and 6-months, 1-year, 2-years, and 3-postoperatively); and, 3) LCG modeling to assess the heterogeneity of the recovery patterns for different subgroups of patients (i.e., age groups and sex).

Methods

The Surgical Outcomes System (SOS; Arthrex, Inc., Naples, FL) was retrospectively queried between the years 2014-2020 to establish a large data set for assessment. Institutional Review Board (IRB) was not required for this study as the SOS is an international registry of de-identified data that adheres to the Health Insurance Portability and Accountability Act (HIPPA) regulations. The SOS contains information submitted from multiple orthopedic surgical centers around the world. This study queried all patients who underwent a THA procedure and completed the 6-item HOOS-JR. Demographic information (i.e., sex, age at treatment) and responses to the 6-item HOOS-JR were downloaded from the SOS for analysis.

Hip Disability and Osteoarthritis Outcome Scale Joint Replacement

The modified HOOS-JR includes five items that ask participants to rate how frequently they engaged in behaviors over the past week using a 5-point Likert scale (0 = none, 1 = mild, 2 = moderate, 3 = severe, and 4 = extreme).[26] The five items included are items two through six of the original 6-item HOOS-JR. To be consistent with the original HOOS-JR, scores of the 5-item HOOS-JR were calculated by creating a raw sum score of the five items (i.e., items two through six of the original scale). These methods are similar to the calculation guidelines presented by Lyman et al.,[42] however, no interval score was calculated. Therefore, the highest score (i.e., 20) indicated total hip disability and the lowest score (i.e., 0) indicated perfect hip health.

Data Analysis

Using Microsoft® Excel for Mac (Version 16.46; Redmon, WA), data were exported from the SOS database and uploaded to Statistical Package for Social Sciences Version 24.0 (IBM Corp., Armonk, NY) for data cleaning and analyses purposes. Because the primary purpose was to assess the HOOS-JR, individuals with missing demographic data were not excluded from analysis and were left as missing values. Univariate outliers were assessed using z-scores, and participants with z-scores that exceeded the cut-off value of |3.3| were removed. The presence of multivariate outliers were also assessed; participants were assessed, flagged and removed if Mahalanobis distance, using a chi-square table with degrees of freedom and p-value = 0.01[31,43] was exceeded.[31] For longitudinal invariance, participants who did not respond to the HOOS-JR items at all five time points were not used in the analysis.

Scale Structure – Confirmatory Factor Analysis

The full sample was used to conduct a CFA to assess the scale structure of the 5-item HOOS-JR using Analysis of Moment Structures (AMOS) software (IBM Corp., Armonk, NY) at each time point (i.e., preoperatively, and 6-months, 1-year, 2-years, and 3-years postoperatively). The modified HOOS-JR was specified as a one-factor, 5-item model to remain consistent with our previous work. Full Information Maximum Likelihood estimation was used to generate the parameter estimates[31,44]; model fit statistics were assessed based on *a priori* values. Goodness-of-fit indices used included the likelihood ratio statistic (CMIN), Goodness of Fit Index (GFI; ≥ 0.95), Comparative Fit Index (CFI; ≥

0.95), Tucker-Lewis Index (TLI; ≥ 0.95), Bollen's Incremental Fit Index (IFI; ≥ 0.95), and Root Mean Square Error of Approximation (RMSEA; ≤ 0.05), and Standardized Root Mean Square Residual (SRMR; ≤ 0.08).[31,44,45] Interpretability, size, and significance of the model's parameter estimates (i.e., factor variances, covariances, and indicator errors) were examined to identify any localized areas of strain in addition to assessing the overall goodness-of-fit.[46]

Longitudinal Invariance Testing

The same criteria utilized for the CFAs were used to assess fit for the invariance model.[31,46] Invariance testing was conducted with the full sample to assess measurement invariance of the 5-item HOOS-JR across five time points (i.e., longitudinal invariance). Longitudinal invariance testing assesses model fit of all time points simultaneously, while the prior step (i.e., assessing the scale structure using CFA) tests each time point individually.[31,46] Therefore, individuals who completed the 6-item HOOS-JR at all five time points (i.e., preoperatively, and 6-months, 1-year, 2-years, and 3-postoperatively) were used to assess invariance across time. A CFI difference (CFI_{DIFF}) of less than 0.01, and the chi-square (χ^2_{DIFF}) with a *p*-value cut-off of > 0.01, was evaluated for structural invariance.[31,46] As the χ^2_{DIFF} test is sensitive to sample size, the CFI_{DIFF} test held a greater influence in decisions regarding invariance testing.[31,46] If a model passed the CFI_{DIFF} test, but exceeded the χ^2_{DIFF} test, invariance testing would continue.

Longitudinal invariance implies patients, across repeated measures, interpret the questions and construct (i.e., hip disability) in the same way.[31,46] If a model is invariant, comparisons across repeated measures are possible which can provide clinicians the ability of assess change over time and conclude the measured change was true and not due to measurement error. As such, configural, metric, and scalar invariance testing was performed to assess the scale structure across five time points.[35,44,46] In the configural invariance model, the latent structure was constrained to be equal across time. Secondly, in the metric invariance model, additional constraints were placed on the item loadings across time.[44,46] If the model met metric invariance requirements, equal variances (i.e., differences in scores at different time points) across time were then assessed.[44,46] Lastly, in the scalar invariance model, the latent structure, item loadings, and item intercepts were set to be constrained (i.e., equal) across time.[44,46] If the model met scalar invariance requirements, equal mean models (i.e., score differences) were tested across time points.[44] If model invariance was not found, investigation of the source leading to noninvariance would be explored by sequentially releasing item loadings or intercepts to explore if a partially invariant model could be identified.[47]

Latent Growth-Curve Modeling

As the model fit of the modified 5-item HOOS held during longitudinal invariance testing, the sample was then subjected to LGC modeling using AMOS.[31] The LGC model analysis allows for the evaluation of model adequacy using model fit indices and model selection criteria, which accounts for measurement error using latent repeated measures.[38,48] As such, the LGC modeling was conducted on the modified 5-item HOOS-JR using Full Information Maximum Likelihood as this method is best for handing missing data.[31] First, an *a priori* was hypothesize to be linear,[44] meaning a decrease in pain and an increase function from preoperative to postoperative. Same goodness-of-fit criteria utilized for the CFAs were used to assess fit for the LGC model.[31,35,44,46] A growth trajectory (i.e., intercept and slope) was assessed for intraindividual and interindividual differences to determine the direction and extent to which the patient's self-perceived hip disability changed from preoperative to 3-years postoperatively. The intercept factor represents the starting point (i.e., the preoperative) for the trajectory of a factor (i.e., 5-item HOOS-JR), whereas the slope represents the change in the trajectory of the 5-item HOOS-JR over time.[49]

The model included two growth parameters: a) the intercept parameter representing an individual's score on the HOOS-JR at preoperative (time point 1), and b) the slope parameter representing the individual's rate of change over the 3-year follow-up.[44] Also, covariances between the intercept and slope factors were also included; this provides an indication of whether patients who started at a lower or higher score for the outcome changed at lower or higher rate.[44] The values assigned to the slope parameters were represented as preoperative = 0, 6-months postoperative = 0.5, 1-year postoperative = 1, 2-years postoperative = 2, and 3-years postoperative = 3.[44,50] If the interindividual growth trajectory was statistically significant (i.e., the sample was heterogeneous), multi-group testing were then conducted to determine if the sample could be further explained.[31,44] Groups of interest that were extracted from the database included preoperative patient characteristics, such as age groups and sex. If a nonlinear growth model was found, the slope parameters would be freely estimated to explore the shape of the growth model.[44,50]

Results

A total of 1707 complete participant responses (i.e., all items of the HOOS-JR were answered at all five time points) were extracted for data cleaning. From the total sample with responses at all five time points, univariate (n = 0; 0.0%) and multivariate (n = 40, 2.3%) outliers were identified at preoperative and were removed during the data cleaning process. Deleted cases included females (N = 23) and males (N = 14), with an average age of 65.05 ± 10.01y (range = 46-86). The final sample (N

= 1667) used for analysis consisted of 48.8% males (N = 769) and 51.2% females (N = 806) with a mean age of 61.72 ± 9.90y (range = 24-90y). Of the final sample of responses, 5.5% (N = 92) were missing a response to sex.

Scale Structure

The CFA of the modified 5-item HOOS-JR at each time point (e.g., preoperatively, 6-months postoperatively) indicated acceptable fit of the data (Table 1). Model fit indices met recommended values for the CFI (range = 0.964-0.982), IFI (range = 0.965-0.986), and SRMR (0.021-0.035). However, TLI (range = 0.929-0.972) and RMSEA (range =0.081-0.126) values slightly exceeded recommendations. Factor loadings were significantly different (p-values < 0.001) and ranged from 0.68-0.82. At time point 3 and 4, modification indices revealed meaningful high error-term covariances between item five (i.e., lying in bed [turning over, maintaining hip position]) and item six (i.e., sitting). However, no cross-loadings were identified between items at any time point.

Longitudinal Invariance Testing of the Alternate HOOS-JR

Analysis of the five time points (i.e., preoperatively, and 6-months, 1-year, 2-years, and 3-postoperatively) revealed the initial model (i.e., equal form) met all model fit indices (CFI = .975; χ^2 = 804.79; RMSEA = 0.040; Table 4). The metric model (i.e., equal loadings) passed both the CFI_{DIFF} test (CFI = 0.969) and the χ^2_{DIFF} test (χ^2 = 145.94). Because the metric model was invariant between time points, examination of the equal latent variable factors was warranted. The equal factor variance model slightly exceeded the CFI_{DIFF} test (CFI = 0.964) and the χ^2_{DIFF} test (χ^2 = 286.66), indicating variances of the latent variables were not equal across time points. The scalar model (i.e., equal intercepts) also slightly exceeded the CFI_{DIFF} test (CFI = 0.946) and the χ^2_{DIFF} test (χ^2 = 702.67), suggesting item-level bias. However, because the scalar model only slightly exceeded the CFI_{DIFF} (i.e., 0.012) and χ^2_{DIFF} test, evaluation of the equal latent means model continued. Upon assessment of the equal latent means, the model did not pass either the CFI_{DIFF} test (CFI = 0.835) or the χ^2_{DIFF} test (χ^2 = 3323.33). When the means were not constrained to be equal across time points, significant mean score improvement was found between time point 1 and time point 2 (i.e., -1.69), time point 2 and time point 3 (-1.75), and time point 3 to time point 4 (i.e., -1.83). The mean scores remained consistent (i.e., -1.83) from time point 4 to time point 5, indicating similar mean scores between the two time points that were lower (i.e., indicating improved hip health maintained across time) when compared to preoperative, 6-months postoperative, and 1-year postoperative mean scores.

Because the model did not meet the strictest criteria for scalar invariance, partial invariance at the scalar level was explored by sequentially releasing intercepts of the items.[31,44,47] Item five (i.e., lying in bed [turning over, maintaining hip position]) was identified as the source of noninvariance; when released, the scalar model (i.e., equal intercepts) met the CFI_{DIFF} test (CFI = .965) and the χ^2_{DIFF} test (χ^2 = 1073.83; Table 5). Because the scalar passed the CFI_{DIFF} and the χ^2_{DIFF} test, evaluation of the equal latent means model continued. Upon assessment of the equal latent means, the model did not pass either the CFI_{DIFF} test (CFI = 0.842) or the χ^2_{DIFF} test (χ^2 = 3962.49; Table 5). When the means were not constrained to be equal across time points, the mean scores improved between time point 1 and time point 2 (-1.66), time point 2 and time point 3 (-1.76), time point 3 to time point 4 (-1.84) and remained consistent between time point 4 and time point 5 (-1.84).

Latent Growth-Curve Modeling

The linear LGC model did not meet the recommended model fit indices (CFI = 0.049, TLI = 0.049, IFI = 0.049, RMSEA = 0.424; Figure 3).[44] These findings suggest that patients responses to the questionnaire are nonlinear. Therefore, exploratory methods were used to identify the shape of the growth trajectory. Slope parameters for time points 1 and 2 were constrained to 0 and 0.5 respectively; however, the remainder of the time points (i.e., 3-5) were freely estimated.[44,50] Upon assessment of these findings, the final slope parameters were defined as follows: preoperative = 0.0, 6-months postoperative = 0.92, 1-year postoperative = 0.96, 2-years postoperative = 1.0, and 3-years postoperative = 1.0. This model met the recommended model fit indices (CFI = 0.986; TLI = 0.986; IFI = 0.986; RMSEA = 0.051; Figure 4). Upon assessment of the means, the estimates pertaining to the intercept and shape were statistically significant ($p < 0.001$). In addition, findings revealed the average mean score for modified 5-item HOOS-JR was 10.48 points at preoperative and decreased over three years (i.e., -8.48 points). When assessing the covariance between the intercept and shape, a negative estimate was identified (i.e., -2.08; Figure 4). Variances of the model estimates were not significantly different for the intercept ($p = 0.06$), however, were statistically different for the shape (i.e., $p = 0.05$).

Model fit indices were also assessed between groups (i.e., sex and age) to determine the differences sex and age groups have on the mean scores and growth trajectories over time. When assessing the LGC model pertaining to sex, the model met all recommended fit indices (CFI = 0.981, TLI = 0.981, IFI = 0.981, RMSEA = 0.043). Upon assessment of the estimates pertaining to the means, the parameters for both the intercept and the shape were statistically significant ($p < 0.05$). Upon assessment of the mean scores, the average mean score for the modified 5-item HOOS-JR at preoperative visit were higher in patients within the female group (10.92) compared to patients in the male group (10.01). In addition, patients in both groups improved their scores over the 3-years, with patients in the female group having an overall greater decrease in scores (i.e., females = -8.98, males

= -7.99). When assessing the covariance between the intercept and shape, a negative estimate was identified for patients in the male group (i.e., -3.72); however, the female group had a small, yet positive change (i.e., 0.32). Lastly, when assessing the variances of the model, estimates related to the intercept (p = 0.050) and shape (p = 0.049) were significant.

When assessing the LGC model for age groups, the model met all recommended fit indices (CFI = 0.979, TLI = 0.979, IFI = 0.973, RMSEA = 0.028). On average, patients in the 65-74 age group had an overall worst score at the preoperative visit (i.e., 11.21) when compared to patients in the other age groups. In addition, patients in all groups increased their scores over the 3-years, with patients in the < 44 age group having the greatest decrease in scores (i.e., -9.30). When assessing the covariance between the intercept and shape, those aged < 45 had the largest negative estimate (-13.09) compared to the other age groups. Lastly, when assessing the variances of the model, estimates related to the intercept and shape were not significant (p = 0.185, p = 0.113, respectively).

Discussion

In previous work by E.N. Miley et al. (unpublished data, 2023), the original 6-item HOOS-JR did not meet the recommended goodness-of-fit indices pertaining to structural validity; therefore, further assessment of the modified 5-item HOOS-JR was recommended. The purpose of our study was threefold: 1) assess structural properties, 2) assess longitudinal invariance properties, and 3) assess LGC model characteristics (i.e., rate of perceive change in scores) of the modified 5-item HOOS-JR in a sample of patients who underwent a THA and completed the original HOOS-JR items at multiple visits (i.e., preoperatively, and 6-months, 1-year, 2-years, and 3-years postoperatively). In addition, we sought to assess the heterogeneity in the responses to the 5-item HOOS-JR across subgroups (i.e., age groups and sex) using LGC modeling to determine is differences exist in recovery following a THA. Our results indicate the modified 5-item HOOS-JR can be used across different time points.

However, caution is warranted when attempting to make comparisons using the scores across time. Therefore, as concerns pertaining to the scale (i.e., high error-term covariances and lack of full measurement invariance) still exist, consideration to implement an instrument that better captures the full scope of recovery pertaining to the hip joint is warranted. We did not identify significant group differences between age groups or sex, indicating regardless of the group (i.e., different age groups or males vs. females), patients answer the modified 5-item HOOS-JR similarly. Lastly, most of the improvement in scores (i.e., patients reporting less hip disability) occurred within the first 6-months to 1-year postoperatively.

Scale Structure – Confirmatory Factor Analysis

Psychometric properties pertaining to structural validity of the modified 5-item HOOS-JR was assessed individually at five time points (i.e., preoperatively, and 6-months, 1-year, 2-years, and 3-postoperatively). Most model fit indices were met at each time point, indicating that the modified 5-item HOOS-JR had suitable structural validity for measuring hip disability at the five time points. However, upon further assessment, time points 3 (i.e., 1-year postoperatively) and time point 4 (i.e., 2-years postoperatively) indicated meaningful high error-term covariances between items five (i.e., lying in bed [turning over, maintaining hip position]) and item six (i.e., sitting). Meaningful error-term covariances indicates possible model misspecification which may demonstrate the need for further refinement of the scale.[46] In addition, we found similar findings within our previous work when assessing scale validity in a larger sample of patients (n = 12215) at one time point (i.e., preoperative); meaningful high error-term covariances were also present between these two items. Overall findings support the need for further refinement of the items, such as re-writing or removing items, to produce a more parsimonious and psychometrically sound instrument.[31,44]

Assessment of the RSMEA value revealed a higher than recommended cut-off value (i.e., ≤ 0.05) at each time point. In our previous work, we identified similar RMSEA findings (i.e., 0.143).[31,44] However due to the small *df* at each individual time point (i.e., 5), we assessed the SRMR in conjunction to the RMSEA as previous studies report that models with small *df* can negatively influence the RMSEA and cause potential model misfit.[51,52] As such, authors recommend to report the SRMR in models containing small *df*[51,52]*;* therefore, SRMR and CFI combined held greater weight in our decision regarding model fit interpretation.[51] With this in mind, SRMR fit-indices were met at all time points.

Longitudinal Invariance Testing

Longitudinal invariance was conducted to determine if the modified 5-item HOOS-JR was structurally invariant across multiple time points. Longitudinal invariance was identified for the configural (equal form) and metric (equal loadings) model, indicating that patients interpret the questions and underlying construct (i.e., hip disability) in the same way across repeated visits.[31,46] However, the scalar model (equal intercepts) slightly exceeded the strictest invariance criterion; therefore, subsequent analyses were performed to determine the item intercept that was the source of invariance.[44,47] Results of these analyses revealed that item five (i.e., lying in bed [turning over, maintaining hip position]) was not interpreted similarly by patients across the repeated visits. When the intercept associated with item five was not constrained, the model passed scalar invariance. These findings allow clinicians to be able to assess change over the repeated visits and allow us to conclude that the change measured was not due to measurement error, but true change in the patients perceived hip disability. As scalar invariance was identified with the release of item five, caution is warranted clinically when incorporating the score of this item in the assessment of mean scores over time.[31,47]

The invariant solution allowed us to assess the equal factor variances and equal means model. Equal latent variances model slightly exceeded the χ^2_{DIFF}vtest and the CFI_{DIFF} difference test, indicating variances were not equal across time. When the variances were not constrained to be equal, there was more variability (i.e., 0.49) pertaining to hip disability at the preoperative compared to the postoperative time points (range = 0.25-0.30), indicating that the variances in scores decreased over time. Upon assessment of the equal latent means model, the χ^2_{DIFF} test and the CFI_{DIFF} difference test were exceeded. When the means were not constrained to be equal, patients responded with the highest score (i.e., more hip disability) preoperatively, and the lowest scores (i.e., less hip disability) at 2- and 3-years postoperative. However, the majority of improvement in scores happened between preoperative and 6-months postoperatively; these findings are similar to previous researchers who noted the most improvement in HOOS-JR scores within the first 6-months postoperative THA.[39-41,53]

Latent Growth-Curve Model

Minimal research has focused on LGC modeling pertaining to outcomes related to THA.[39-41] Prior studies pertaining to the OHS,[39,40] HOOS Physical Function (PS),[41] and HHS[53] have included LGC modeling. To our knowledge, this was the first study to assess LGC modeling in patients who answered questions to the HOOS-JR over any longitudinal period, and between age and sex groups following THA. Utilizing LGC modeling is unique because it allows researchers to assess individual growth and interindividual differences in change longitudinally.[44,48,54]

Our LGC model did not meet the recommended model fit indices pertaining to a purely linear model (i.e., constant rate of change over time). However, when the intercepts were constrained to be nonlinear, the model met recommended model fit indices. These findings suggest that patients improve their hip disability, as measured by the modified 5-item HOOS-JR, at a nonlinear rate. With the use of LGC modeling, we identified significant differences for the intercept and shape trajectory ($p < 0.001$). Overall, patients average mean scores on the modified 5-item HOOS-JR were 10.48 and decreased by 8.48 points from preoperative to 3-years postoperative. These findings suggest that patients decrease their scores on the 5-item HOOS-JR over time (i.e., improved hip disability) following a THA. Within our model, most score improvement occurred within the first 6-months postoperatively. Also, when assessing the covariance between intercept and shape trajectory, a negative value was identified (-2.08), indicating that patients who reported less hip disability at the preoperative visit demonstrated a lower rate of improvement in HOOS-JR scores over the 3-years postoperatively.[31,44] These findings are consistent with prior THA research that identified the largest improvement in PRO scores (i.e., OHS, HHS, HOOS-PS) within the first 3- to 6-months postoperative.[39-41,53] In addition, changes in mean scores of the modified 5-item HOOS-JR is similar to other studies assessing the differences in PRO scores over time.[39-41,53]

We also assessed between group (i.e., age groups and sex) differences to test for differences between sex and age groups across time. When assessing the means (i.e., intraindividual differences), the intercept and shape trajectory were statistically significant ($p < 0.001$). These findings reveal that the average scores of the modified 5-item HOOS-JR notably differ; average scores were lower for males (i.e., 10.01) compared to females (10.92). As such, patients in the female group reported more hip disability compared to the male patients at preoperative visits. Our findings are consistent with previous literature that indicated females reported higher levels of hip disability than males prior to surgery.[55,56] In addition, patients in both the male and female groups demonstrated a decrease in average scores (-7.99 and -8.98, respectively) from preoperatively to 3-years postoperatively.

When assessing the covariances between intercept and slope, females had a faster change in mean scores (i.e., 0.32) compared to males (i.e., -3.73); however, the negative covariance represents patients who scored lower at preoperative (i.e., less hip disability) demonstrated a lower rate of increase in scores over the 3-year period. Lastly, when assessing variances of the model (i.e., interindividual differences), estimates related to the intercept ($p = 0.050$) and shape ($p = 0.049$) were significant. These findings reveal that there are significant differences in the variances of the mean scores at the preoperative visit and over time associated with the modified 5-item HOOS-JR. To our knowledge, this was the first study assessing the variances in scores between males and females longitudinally (i.e., over time) from preoperatively to 3-years postoperatively. In a previous study, authors, report the mean score differences at preoperative and 5-years postoperatively.[53] Even though the authors found similar findings at the preoperative visit (i.e., significant differences between males and females), the amount of hip disability was not significantly different between sexes at the 5-years postoperative visit.[53]

In addition, we sought to determine if differences in scores existed between age groups (i.e., < 45, 45-54, 55-64, 65-74, ≥ 75). Our LGC model met all recommended fit indices (CFI = 0.98, TLI = 0.978, IFI = 0.980, RMSEA = 0.029). When assessing the means, the intercept and the shape trajectory were statistically significant ($p < 0.001$); patients in the < 45 age group reported significantly higher scores of the modified 5-item HOOS-JR compared to patients within the other age groups. In addition, patients in the 65-74 age group reported the lowest average score (i.e., better hip health) on the modified 5-item HOOS-JR scores (10.08) preoperatively. However, the average scores significantly decreased the most for patients within the < 45 age groups (-9.30), whereas patients in the 65-74 age group decreased the least (-8.15). Our findings reveal that those who are younger reported greater improvement in their hip disability from preoperative to 3-years postoperatively compared to other age groups.[44] Our findings support prior research that demonstrates younger patients report faster recovery patterns following a THA when compared to older patients.[40]

Lastly, we identified a larger negative covariance between intercept and slope (i.e., -13.09) in patients in the < 45 age group. These findings reveal that patients < 45 who reported lower scores (i.e., less hip disability) of the modified 5-item HOOS preoperatively demonstrate a lower rate of improvement in scores over the 3-years postoperatively; whereas patients in the 55-64 age group had a larger, positive (i.e., 2.89) covariance, suggesting patients have a faster rate of improvement in hip disability as measured by the modified 5-item HOOS-JR scores. Of note, however, even though we identified intraindividual differences in average mean scores between age groups, these findings were not statistically different between groups. When assessing interindividual differences in both the initial scores and their change over time (i.e., variances of the model), estimates pertaining to the intercept ($p = 0.185$) and slope ($p = 0.113$) were not significant.

Limitations and Future Research

Even though this study is the first to our knowledge to explore longitudinal invariance testing and LGC modeling of the HOOS-JR, several limitations were present warranting discussion. Even though we had a large sample of patients who completed the 6-item HOOS-JR at five different time points, further research should include a confirmation sample to ensure similar findings exist of the modified 5-item HOOS-JR. Our sample lacked complete demographic information pertaining to the dataset limiting further understanding of the population assessed. In addition, the lack of demographic information restricted the analyses of additional subgroups (e.g., surgical approach, ethnicity, race, socioeconomic status). Future research should include more demographic information for further testing. Clinicians should practice caution when examining group differences in scores of the modified 5-item HOOS-JR in subgroups not yet analyzed. We performed longitudinal invariance testing and LGC modeling on the modified 5-item HOOS-JR which was identified in our previous work. However, the sample of patients responded to the 6-item HOOS-JR. It is possible responses were influenced by the additional item present on the original 6-item HOOS-JR scale. Future research should be conducted on patients who only respond to the items on the modified scale. We found item five to be problematic (i.e., high error-term covariance); as such, partial invariance was identified for the structure by removing the intercept constraint associated with the item during assessment of the scalar model. Therefore, further scale modifications (e.g., re-word or removal of the item) may be pertinent and caution is recommended when assessing the change in mean scores of the 5-item HOOS-JR across time.

Conclusion

To our knowledge, this was the first study to assess longitudinal invariance testing and perform multi-group LGC modeling approach for the modified 5-item HOOS-JR. As the 5-item HOOS-JR did not meet the strictest longitudinal measurement invariance criteria (i.e., full scalar invariance), partial scalar invariance was identified by releasing the scores pertaining to item five. In addition, we identified that patients self-report most improvement in their scores within the first 6-months postoperatively. We found that significant differences in mean scores between sex, with females reporting more hip disability preoperatively and report a faster improvement as measured by scores of the modified 5-item HOOS-JR. Researchers and clinicians can use the scale at different time points; however, caution is warranted when attempting to compare mean scores across time.

References

1. Gould D, Thuraisingam S, Shadbolt C, et al. Cohort profile: the St Vincent's Melbourne Arthroplasty Outcomes (SMART) Registry, a pragmatic prospective database defining outcomes in total hip and knee replacement patients. *BMJ open.* 2021;11(1):e040408.
2. Ethgen O, Bruyere O, Richy F, Dardennes C, Reginster J-Y. Health-related quality of life in total hip and total knee arthroplasty: a qualitative and systematic review of the literature. *J Bone Joint Surg.* 2004;86(5):963-974.
3. Koenig L, Zhang Q, Austin MS, et al. Estimating the societal benefits of THA after accounting for work status and productivity: a Markov model approach. *Clinical Orthop Relat Res.* 2016;474(12):2645-2654.
4. Trudelle-Jackson E, Emerson R, Smith S. Outcomes of total hip arthroplasty: a study of patients one year postsurgery. *J Orthop Sports Phys Ther.* 2002;32(6):260-267.
5. Ackerman IN, Bohensky MA, Zomer E, et al. The projected burden of primary total knee and hip replacement for osteoarthritis in Australia to the year 2030. *BMC Musculoskelet Disord.* 2019;20(1):1-10.
6. Kurtz S, Ong K, Lau E, Mowat F, Halpern M. Projections of primary and revision hip and knee arthroplasty in the United States from 2005 to 2030. *JBJS.* 2007;89(4):780-785.
7. Crawford RW, Murray D. Total hip replacement: indications for surgery and risk factors for failure. *Ann Rheum Dis.* 1997;56(8):455-457.
8. DeShazo JP, Hoffman MA. A comparison of a multistate inpatient EHR database to the HCUP Nationwide Inpatient Sample. *BMC Health Serv Res.* 2015;15(1):1-8.
9. Quality. Overview of the National (Nationwide) Inpatient Sample (NIS). Published December 2020. Accessed March 20, 2021.
10. Hash-Campbell D. *Predicting Functional Outcomes Post Total Hip Arthroplasty in Women Under 50*, Weill Medical College of Cornell University; 2020.
11. Zhang Y, Jordan JM. Epidemiology of osteoarthritis. *Clin Geriatr Med.* 2010;26(3):355-369.
12. Sinusas K. Osteoarthritis: diagnosis and treatment. *Am Fam Physician.* 2012;85(1).

13. Clair SFS, Higuera C, Krebs V, Tadross NA, Dumpe J, Barsoum WK. Hip and knee arthroplasty in the geriatric population. *Clin Geriatr Med.* 2006;22(3):515-533.

14. Livermore-Brasher JS, Baker R, Nasypany A, Cheatham SW. Guidelines for Useful Integration of Patient-Rated Outcome Measures into Clinical Practice. *Clinical Practice in Athletic Training.* 2018;1(2):14-31.

15. Bachmeier CJ, March L, Cross M, et al. A comparison of outcomes in osteoarthritis patients undergoing total hip and knee replacement surgery. *Osteoarthritis Cartilage.* 2001;9(2):137-146.

16. Dawson J, Fitzpatrick R, Murray D, Carr A. Comparison of measures to assess outcomes in total hip replacement surgery. *BMJ Qual Saf.* 1996;5(2):81-88.

17. Bulstrode C, Murray D, Carr A, Pynsent P, Carter S. Designer hips. *BMJ.* 1993;306(6880):732.

18. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. 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.* 1988;15(12):1833-1840.

19. Kalairajah Y, Azurza K, Hulme C, Molloy S, Drabu KJ. Health outcome measures in the evaluation of total hip arthroplasties—a comparison between the Harris hip score and the Oxford hip score. *J Arthroplasty.* 2005;20(8):1037-1041.

20. Lyman S, Lee YY, McLawhorn AS, Islam W, MacLean CH. What are the minimal and substantial improvements in the HOOS and KOOS and JR versions after total joint replacement? *Clinical Orthop Relat Res.* 2018;476(12):2432.

21. Matharu G, McBryde C, Robb C, Pynsent P. An analysis of Oxford hip and knee scores following primary hip and knee replacement performed at a specialist centre. *Bone Joint J.* 2014;96(7):928-935.

22. Nilsdotter A, 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. *Arthrit Care Res.* 2011;63(S11):S200-S207.

23. Nilsdotter AK, Lohmander LS, Klässbo M, Roos EM. Hip disability and osteoarthritis outcome score (HOOS)–validity and responsiveness in total hip replacement. *BMC Musculoskelet Disord.* 2003;4(1):10.

24. Ostendorf M, Van Stel H, Buskens E, et al. Patient-reported outcome in total hip replacement: a comparison of five instruments of health status. *J Bone Joint Surg Br.* 2004;86(6):801-808.

25. Rahman WA, Greidanus NV, Siegmeth A, Masri BA, Duncan CP, Garbuz DS. Patients report improvement in quality of life and satisfaction after hip resurfacing arthroplasty. *Clinical Orthop Relat Res.* 2013;471(2):444-453.

26. Lyman S, Lee YY, Franklin PD, Li W, Mayman DJ, Padgett DE. Validation of the HOOS, JR: a short-form hip replacement survey. *Clinical Orthop Relat Res.* 2016;474(6):1472-1482.

27. Gandek B, Roos EM, Franklin PD, Ware Jr JE. A 12-item short form of the Hip disability and Osteoarthritis Outcome Score (HOOS-12): tests of reliability, validity and responsiveness. *Osteoarthritis Cartilage.* 2019;27(5):754-761.

28. Leech NL, Barrett KC, Morgan GA. *IBM SPSS for intermediate statistics: Use and interpretation.* Routledge; 2014.

29. Taber KS. The use of Cronbach’s alpha when developing and reporting research instruments in science education. *Research in Science Education.* 2018;48(6):1273-1296.

30. Burton LJ, Mazerolle SM. Survey instrument validity part I: Principles of survey instrument development and validation in athletic training education research. *Athletic Training Education Journal.* 2011;6(1):27-35.

31. Kline RB. *Principles and practice of structural equation modeling.* Guilford publications; 2015.

32. Matsunaga M. How to Factor-Analyze Your Data Right: Do's, Don'ts, and How-To's. *Int J Psychol Res.* 2010;3(1):97-110.

33. Miley EN, Casanova MP, Cheatham SW, Larkins L, Pickering MA, Baker RT. Confirmatory Factor Analysis of the Hip Disability and Osteoarthritis Outcome Score (HOOS) and Associated Sub-scales. *Int J Sports Phys Ther.* 2023.

34. Bryant FB, Yarnold PR. Principal-components analysis and exploratory and confirmatory factor analysis. 1995.

35. Byrne BM. *Structural equation modeling with Mplus: Basic concepts, applications, and programming.* Routledge; 2013.

36. Singer JD, Willett JB, Willett JB. *Applied longitudinal data analysis: Modeling change and event occurrence.* Oxford university press; 2003.

37. Bolander W, Dugan R, Jones E. Time, change, and longitudinally emergent conditions: understanding and applying longitudinal growth modeling in sales research. *Journal of Personal Selling and Sales Management.* 2017;37(2):153-159.

38. Preacher KJ, Wichman AL, MacCallum RC, Briggs NE. *Latent growth curve modeling.* Sage; 2008.

39. Hesseling B, Mathijssen N, Van Steenbergen L, Melles M, Vehmeijer S, Porsius J. Fast starters, slow starters, and late dippers: trajectories of patient-reported outcomes after total hip arthroplasty: results from a Dutch nationwide database. *J Bone Joint Surg.* 2019;101(24):2175-2186.

40. Porsius JT, Mathijssen NM, Klapwijk-Van Heijningen LC, Van Egmond JC, Melles M, Vehmeijer SB. Early recovery trajectories after fast-track primary total hip arthroplasty: the role of patient characteristics. *Acta Orthop.* 2018;89(6):597-602.

41. Golinelli D, Grassi A, Tedesco D, et al. Patient reported outcomes measures (PROMs) trajectories after elective hip arthroplasty: a latent class and growth mixture analysis. *J Patient Rep Outcomes.* 2022;6(1):95.

42. Lyman S. Hip Disability and Osteoarthritis Outcome Score for Joint Replacement (HOOS, JR.) Scoring Instructions. Hospital for Special Surgery. https://www.hss.edu/hoos-jr-koos-jr-outcomes-surveys.asp. Published 2022. Accessed.

43. Tabachnick BG, Fidell LS. Using multivariate statistics.(4th edn.) Needham Heights. *MA: Allyn & Bacon.* 2001.

44. Byrne BM. Structural equation modeling with AMOS: basic concepts, applications, and programming (multivariate applications series). *New York: Taylor & Francis Group.* 2010;396:7384.

45. Grimm LG, Yarnold PR. *Reading and understanding multivariate statistics.* American Psychological Association; 1995.

46. Brown TA. *Confirmatory factor analysis for applied research.* Guilford publications; 2015.

47. Putnick DL, Bornstein MH. Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental review.* 2016;41:71-90.

48. Grimm KJ, Ram N, Estabrook R. *Growth Modeling: Structural Equation and Multilevel Modeling Approaches.* Guilford Publications; 2016.

49. Mansell G, Hill JC, Main CJ, Von Korff M, Van Der Windt D. Mediators of treatment effect in the back in action trial: using latent growth modeling to take change over time into account. *Clin J Pain.* 2017;33(9):811.

50. Burant CJ. Latent growth curve models: Tracking changes over time. *Int J Aging Hum Dev.* 2016;82(4):336-350.

51. Taasoobshirazi G, Wang S. The performance of the SRMR, RMSEA, CFI, and TLI: An examination of sample size, path size, and degrees of freedom. *Journal of Applied Quantitative Methods.* 2016;11(3):31-39.

52. Shi D, DiStefano C, Maydeu-Olivares A, Lee T. Evaluating SEM model fit with small degrees of freedom. *Multivariate Behav Res.* 2020:1-36.

53. Ng C, Ballantyne J, Brenkel I. Quality of life and functional outcome after primary total hip replacement: a five-year follow-up. *J Bone Joint Surg Br.* 2007;89(7):868-873.

54. Tomarken AJ, Waller NG. Structural equation modeling: Strengths, limitations, and misconceptions. *Annu Rev Clin Psychol.* 2005;1:31-65.

55. Raja A, Jenkins A, Reams M, Horst PK. Normative data of the Hip Disability and Osteoarthritis Outcome Score, JR in a healthy United States population. *J Arthroplasty.* 2019;34(6):1122-1126.

56. Mannion AF, Impellizzeri FM, Naal FD, Leunig M. Women demonstrate more pain and worse function before THA but comparable results 12 months after surgery. *Clinical Orthop Relat Res.* 2015;473:3849-3857.

Tables

Table 4

Goodness-of-Fit Indices for Longitudinal Invariance Across Time Points

Modified 5-item HOOS-JR	χ^2	***df***	χ^2_{diff} **(df_{diff})**	**CFI**	**CFI_{diff}**	**TLI**	**RMSEA**	**SRMR**
Preoperative	65.33	5	----	0.982	----	0.965	0.085	0.024
6-month postoperative	59.84	5	----	0.982	----	0.964	0.081	0.025
1-year postoperative	94.98	5	----	0.974	----	0.949	0.104	0.028
2-year postoperative	138.08	5	----	0.964	----	0.929	0.126	0.035
3-year postoperative	63.30	5	----	0.986	----	0.972	0.084	0.021
Configural (equal form)	804.79	216	----	0.975	----	0.965	0.040	0.028
Metric (equal loadings)	950.73	232	145.94 (16)	0.969	0.006	0.960	0.043	0.033
Equal factor variances*	1091.45	236	**286.66 (20)**	0.964	**0.011**	0.952	0.047	0.059
Scalar (equal indicator intercepts)	1507.46	248	**702.67 (32)**	0.946	**0.012**	0.935	0.055	0.033
Equal latent means*	4128.12	252	**3323.33 (36)**	0.835	**0.140**	0.804	0.096	**0.232**

***** = Substantive questions; **Bolded** = did not meet cuff off criteria.

Table 5

Goodness-of-Fit Indices for Partial Longitudinal Invariance Analyses Across Time Points

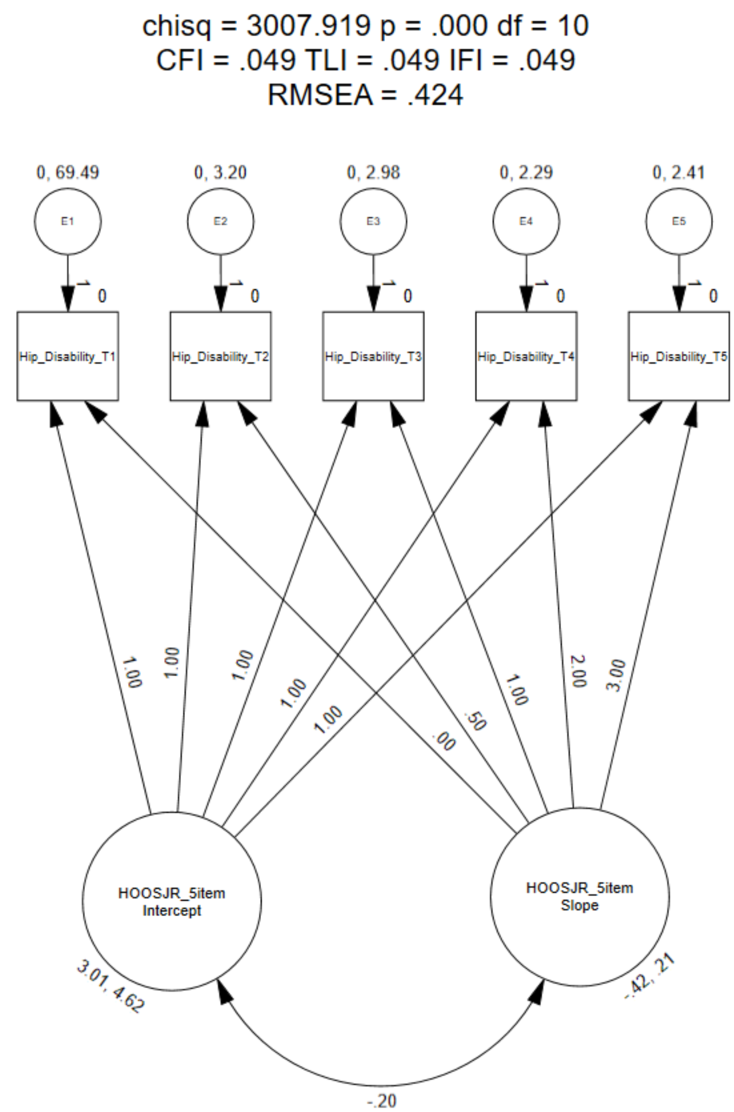
Modified 5-item HOOS-JR	χ^2	*df*	χ^2_{diff} (df_{diff})	**CFI**	CFI_{diff}	**TLI**	**RMSEA**	**SRMR**
Preoperative	65.33	5	----	0.982	----	0.965	0.085	0.024
6-month postoperative	59.84	5	----	0.982	----	0.964	0.081	0.025
1-year postoperative	94.98	5	----	0.974	----	0.949	0.104	0.028
2-year postoperative	138.08	5	----	0.964	----	0.929	0.126	0.035
3-year postoperative	63.30	5	----	0.986	----	0.972	0.084	0.021
Configural (equal form)	804.79	216	----	0.975	----	0.965	0.040	0.028
Metric (equal loadings)	950.73	232	145.94 (16)	0.969	0.006	0.960	0.043	0.033
Equal factor variances*	1091.45	236	**286.66 (20)**	0.964	**0.011**	0.952	0.047	0.059
Scalar (equal indicator intercepts)**	1073.83	244	269.04 (28)	0.965	0.010	0.957	0.040	0.031
Equal latent means*	3962.49	248	**3157.70 (32)**	0.842	**0.133**	0.809	0.095	**0.262**

***** = Substantive questions; **Bolded** = did not meet cuff off criteria; ** = Release of item 5 intercept at level of invariance testing.

Figures

Figure 3

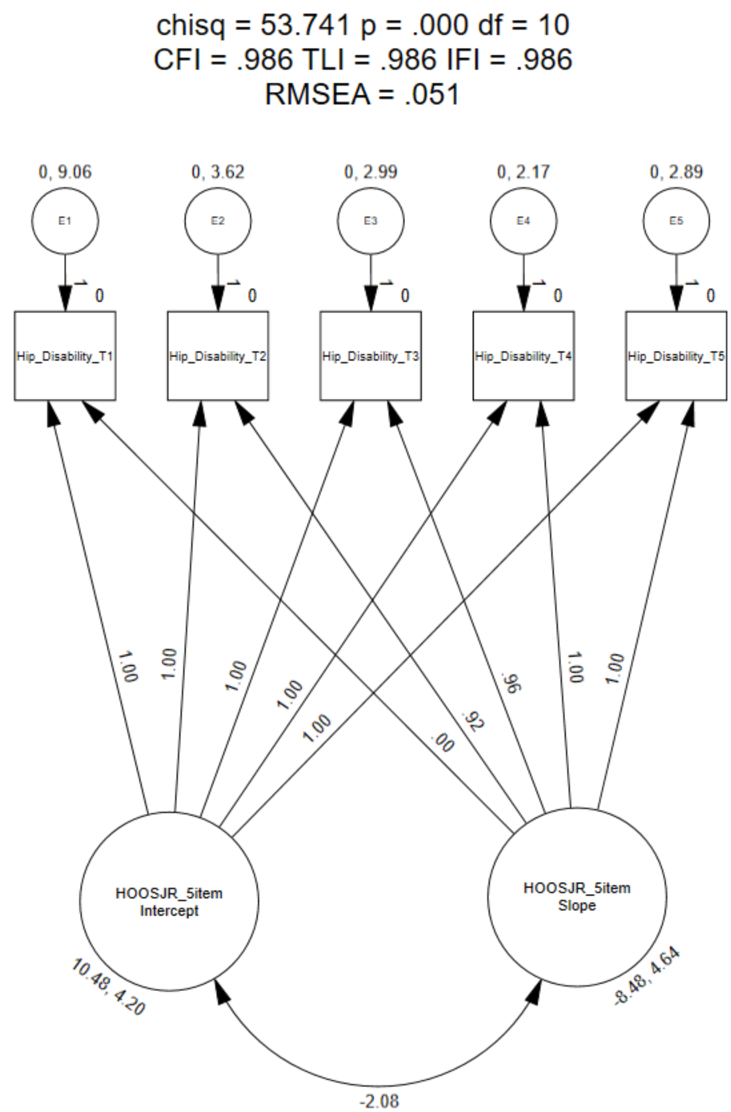
Linear Latent Growth Curve Model of the Modified 5-item HOOS-JR



Chisq = Chi Square (χ^2); df = degrees of freedom, p = alpha level; CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; IFI = Bollen's Incremental Fit Index; RMSEA = Root Mean Square Error of Approximation.

Figure 4

Exploratory Latent Growth-Curve Model of the Modified 5-item HOOS-JR



Chisq = Chi Square (χ^2); df = degrees of freedom, p = alpha level; CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; IFI = Bollen's Incremental Fit Index; RMSEA = Root Mean Square Error of Approximation.

Chapter 3

Manuscript 3: Psychometric Analysis of the Hip Disability and Osteoarthritis Outcome Score (HOOS)

Abstract

Background: The Hip Disability and Osteoarthritis Outcome Survey (HOOS) was developed as a region- and disease specific- outcome to assess hip disability. As such, the primary purpose of this study was to assess the structural validity of the HOOS in patients who underwent a THA. If model fit recommendations for the instrument were not met, alternate model generation procedures would be performed to determine if an appropriate model could be identified from the included items. Secondary purposes of this study were to conduct invariance testing between age groups, sex, and across time points and latent growth-curve (LGC) modeling on the parsimonious scale structure identified.

Methods: Data were obtained from the Surgical Outcome System (SOS) global registry. A confirmatory factor analysis (CFA) was conducted to assess the scale structure of the 40-item HOOS. If model fit indices were not met, the dataset would be randomly split into two samples (i.e., n1, n2). Exploratory factor analysis (EFA) using maximum likelihood extraction with direct oblimin rotation was conducted on one sample (n1) to identify a parsimonious solution of the scale structure.

Results: The original five-factor, 40-item HOOS did not meet recommended model fit indices values (CFI = 0.822, TLI = 0.809, IFI = 0.822, RMSEA = 0.085). Alternate model generation using sample n1 identified an alternative model (i.e., HOOS-9). Sound model fit was found in sample n2 for the HOOS-9 (CFI = 0.974, TLI = 0.961, RMSEA = 0.046). Invariance testing criteria were also met between groups (i.e., age and sex) and across time. Lastly, a nonlinear growth trajectory was identified in responses pertaining to hip disability.

Conclusion: The original scale structure of the 40-item HOOS was not supported. An alternate three-factor, 9-item model (i.e., HOOS-9) was identified that met contemporary model fit recommendations, along with multi-group and longitudinal invariance testing. Our findings support the use of the alternate HOOS-9 as a more viable option to assess hip function and disability in research and clinical practice.

Keywords: total hip arthroplasty, confirmatory factor analysis, invariance testing, latent growth-curve modeling

Introduction

Hip Disability and Osteoarthritis Outcome Survey (HOOS) was developed as a region- and disease-specific outcome to assess disability pertaining to osteoarthritis (OA).[1] The HOOS development process relied heavily on two previously developed instruments: 1) the Western Ontario McMaster Osteoarthritis Score (WOMAC), and 2) the Knee Injury and Osteoarthritis Outcome Score (KOOS). The WOMAC is a disease-specific instrument validated for OA in the lower extremities, and for evaluating outcomes after a total hip arthroplasty (THA)[1,2] while the KOOS is a region-specific instrument intended to measure pain, symptoms activities of daily living (ADLs), sport and recreation function, and knee-related quality of life (QOL) in middle aged patients with or without knee OA.[3] The HOOS contains items (n = 24) and proposed constructs (i.e., pain, stiffness, physical function) of the WOMAC[1,4]; the HOOS also contains items (n = 11) and proposed constructs (i.e., sport and recreational function, and QOL) derived from the KOOS to expand the constructs measured from the WOMAC items and to improve scale sensitivity and responsiveness in younger, more athletically active patients undergoing a THA for treatment of OA.[1] Lastly, the HOOS designers constructed five additional items: two in the pain construct, two in the symptoms construct, and one in the sport/recreation construct.

In total, the HOOS assesses five dimensions of hip-related health with 40 items: Pain (P: 10 items), Symptoms (S: five items), Function limitations-daily living (A: 17 items), Sport and Recreation Function (SP: four items), and Hip Related Quality of life (Q: four items).[1,4] The items are answered using five-option Likert-boxes with three different scale options (i.e., never to always, none to extreme, or never to constantly) across the five constructs.[1,4] All of the items are scored zero to four; each dimension is individually scored then transformed into a 0-100 scale,[1,4] with 100 indicating no symptoms and 0 indicating extreme symptoms.[1,4] Currently, the HOOS has been derived to accommodate 26 languages and is recognized in the United States as an acceptable outcome for measuring functional assessment in patients 21 years of age and older who have been diagnosed with OA.[5] Although the HOOS was developed to assess outcomes in patients with OA, the HOOS is also used globally for reimbursement purposes to assess short-term and long-term changes induced by a variety of treatment options including, but not limited to, THA.[4] As such, it is pertinent to assess the measurement properties of the HOOS across the diverse patient populations in which it is used.

Early scale validation work has focused on validity (i.e., construct), reliability, and instrument responsiveness.[1,6-8] Construct validity of the HOOS was assessed by correlating the constructs to the SF-36: moderate correlations (r = 0.49–0.66) were identified between constructs measuring physical health (i.e., function and pain) using Spearman's correlation coefficient; however, weaker correlations were identified for assessment of mental health ($r = 0.26$).[1] Additionally, Cronbach's alpha values for the constructs (e.g., pain, symptoms) of the HOOS ranged from 0.75 to 0.98 across multiple studies.[6-8] Values ranging from 0.70 to 0.89 have been generally recommended for each construct within an instrument[9-11]: 1) exceptionally high values (i.e., > 0.90) may be indicative of item redundancy, parallel items, construct underrepresentation, inclusion of too many items, and reduced precision[11,12 13,14] and 2) low values (i.e., < 0.70 in general; ≤ 0.80 for research tools) may indicate poor internal consistency within the instrument.[9,10,12-14]

Test-retest reliability of the HOOS has also been reported with values ranging from good to excellent (ICC = 0.75 to 0.97).[6-8] Responsiveness (i.e., the validity of the HOOS over time) has also been established; the HOOS was significantly more responsive in the pain and symptoms constructs (SRM: 2.11, 1.83) compared to the pain and stiffness constructs of the WOMAC (SRM: 1.83, 1.28).[1] Lastly, patients younger than 66 years of age reported a higher responsiveness in all five constructs of the HOOS compared to patients over the age of 66 years.[1]

Few researchers have examined the psychometric properties of the HOOS using exploratory factor analysis (EFA) or confirmatory factor analysis (CFA) and invariance procedures to verify the underlying factor structure and ensure measurement invariance as is recommended in scale development.[8-12] Minimal CFAs have been published on individual constructs proposed in the original HOOS,[15] and recommended practices and findings have not always been reported.[9-12] Previous authors performed CFAs on individual constructs and model fit recommendations were met for pain (i.e., CFI = 0.99, TLI = 0.98) and function (i.e., CFI = 0.97, TLI = 0.97); however, other fit indices recommendations were not met (i.e., RMSEA = 0.14-0.19) for these contructs.[15] The complete scale model (i.e., all constructs) should also be assessed using factor analysis procedures.[9-12] Researchers have reported that the full scale structure did not meet contemporary fit recommendations (CFI = 0.847; TLI = 0.836; IFI = 0.847; RMSEA = 0.098) in a sample of primarily self-reported healthy participants.[16] Further, correlations found between first-order latent constructs (e.g., pain, function) were high (ranging from $r = 0.80$ to $r = 0.96$), modification indices revealed several meaningful cross-loadings were present (e.g., putting on socks/stockings, taking off socks/stockings), and assessment of error term correlations revealed that most of the items shared commonalities.[16] The overall findings do not support the factorial validity of the original HOOS structure and suggest the presence of multicollinearity, overlapping items or items that are perceived to ask similar questions, and reduced measurement precision.[10-12]

The reported poor psychometric properties of the HOOS were not surprising given that the scale is predominantly derived from the WOMAC, which has also been reported to have questioned psychometric properties. For example, poor fit indices values and error-term correlation findings on the HOOS[16] are similar to those found when examining the scale structure of the WOMAC.[17,18] Authors identified that the pain construct was not supported as a single factor with uncorrelated error terms (CFI = 0.90; TLI = 0.80; RMSEA = 0.21), and the modification indices revealed significant error correlations between “at night while in bed” and “sitting or lying”, “walking on flat ground” and “going up or down stairs”.[18] Researchers examining the scale structure of the WOMAC performed a CFA on 11 of the 24 items (i.e., 3 pain items, 8 function items) and reported moderate overall fit in two samples (CFI = 0.954-0.965; RMSEA = 0.069-0.079).[17,17] However, error-term correlations were specified between the items in this model which included: "pain walking on flat surface" and "functional limitation walking on flat surface", "pain up/down stairs" and "functional limitation ascending stairs", and "functional limitation getting in/out of a car" and "functional limitation putting on socks".[17] However, the addition of error-term correlations between items limits the conclusions that can be drawn from the scale; as such, previous methods of scoring may not be sufficient as the items correlated cannot be scored separately.[19] Therefore, difficulties arise when trying to interpret the scoring for the instrument and is not recommended for best practices.[12,19]

While the previous findings call into question the factorial validity of the HOOS and WOMAC, there were limitations noted in the previous studies worth considering. First, the CFAs were generally performed on individual constructs instead of examining the full factor structure as is recommended.[11,12,20,21] Secondly, the one study which included a CFA on the full model utilized a moderate sample (n = 655) of mostly healthy respondents.[16] Further, invariance testing (e.g., multi-group, longitudinal) results have not been reported in the target population (i.e., THA patients) and this testing is an important process that ensures the interpretations between groups (e.g., male vs. females) or across time (e.g., preoperative and postoperative visits) are valid and reliable.[9,22]

Despite the current use of the HOOS (e.g., approved outcome measure for reimbursement purposes following THA), complete and robust psychometric analysis of the scale in a large dataset of patients for which the scale is designed has not been performed. As such, there is a need to conduct a CFA on the full scale model t test the hypothesized factor structure of the 40-item HOOS to ensure that the items are appropriate indirect measures of the hypothesized latent constructs in a large targeted sample of patients seeking care for a hip pathology (e.g., THA).[21,22] If the scale structure fails to meet recommended levels during the CFA procedures, alternate model generation should be conducted, following best practice recommendations,[23] on the given items to determine if a parsimonious scale structure can be identified prior to further testing.[12,20,21] Additionally, there is need for invariance testing to ensure the scale is unbiased towards different groups of identified models which meet recommendations.[12,20,21] Lasty, it is important to understand how different groups respond to the outcomes over time postoperatively.

Therefore, the primary purpose of this study was to assess the psychometric properties of the HOOS in a large, diverse sample of patients who underwent a THA. If model fit recommendations for the instrument were not met, alternate model generation procedures would be performed to determine if an appropriate model could be identified in the included items. Secondary purpose of this study was to conduct invariance testing between age groups and sex (i.e., multi-group), across time points (i.e., longitudinal), and to perform latent growth-curve (LGC) modeling on the parsimonious scale structure identified.

Methods

Data were obtained for analysis from an international patient reported outcomes database (i.e., Surgical Outcome System [SOS]) global registry maintained by Arthrex (Arthrex, Inc., Naples, FL). University Institutional Review Board (IRB) indicated approval for this study was not required because analysis of the deidentified data set from the SOS database was not considered human subject research. However, IRB approval was granted by the Cedar-Sinai Office of Research Compliance and Quality Improvement as part of a larger research project using SOS data. Patients who underwent a THA and completed the HOOS preoperatively and postoperatively were included in the sample. Patient demographics including sex, age at treatment, race, and ethnicity were included with HOOS responses.

Data Analysis

Data were exported from the SOS database using Microsoft® Excel for Mac (Version 16.46; Redmon, WA). Once downloaded, the data were then uploaded to Statistical Package for Social Sciences (SPSS) version 28.0 (IBM Corp., Armonk, NY) for purposes of data cleaning and analyses. As the primary purpose was to assess the 40-item HOOS, individuals with missing demographic data were left as missing values and not excluded from analysis. Univariate outliers were assessed using z-scores, and if they exceeded the |3.3| cut-off value removal was warranted. In the presence of multivariate outliers, participants were assessed, flagged, and removed if Mahalanobis distance was exceeded.[12,24] The cut-off value was identified using a chi-square table with degrees of freedom (df = 40) and p-value < 0.01.[12,24]

Scale Structure - Confirmatory Factor Analysis

Analysis of Moment Structures (AMOS) software version 27 (IBM Corp., Armonk, NY) was used to conduct a CFA to assess the scale structure of the 40-item HOOS. The original HOOS was specified as a five-factor, 40-item model to remain consistent with the original proposed model.[1] To generate the parameter estimates, Full Information Maximum Likelihood estimation was used[12,20]; model fit statistics were assessed based on *a priori* values. Indices used to assess the goodness-of-fit included the likelihood ratio statistic (CMIN), Goodness of Fit Index (GFI; ≥ 0.95), Comparative Fit Index (CFI; ≥ 0.95), Tucker-Lewis Index (TLI; ≥ 0.95), Bollen's Incremental Fit Index (IFI; ≥ 0.95), and Root Mean Square Error of Approximation (RMSEA; ≤ 0.05).[12,20,25] In addition to assessing the overall goodness-of-fit, interpretability, size, and significance of the model's parameter estimates (i.e., factor variances, covariances, and indicator errors) were examined to identify any localized areas of strain.[21]

Alternate Model Generation

The dataset was randomly split into two samples (n1, n2). Exploratory factor analysis using maximum likelihood extraction with direct oblimin rotation was conducted in SPSS on sample n1 to identify a parsimonious solution of the scale structure. Criteria utilized to determine the number of factors retained included: (1) factors with an eigenvalue > 1.0, (2) scree plot inflexion point examination, and (3) factors that accounted for more than 5% of the variance.[9,21,23,26] To confirm the number of factors to obtain, parallel analysis was then conducted.[27]

Bartlett's test for sphericity and Kaiser-Meyer Olkin (KMO) Measure of Sampling Adequacy were both assessed for violations. Cut-off values were set *a priori* at < 0.01 for Bartlett's test of sphericity and ≥ 0.80 for KMO, which are conservative compared to widely accepted values (KMO > 0.70, Bartlett's < 0.05).[9] Items were assessed individually and removed one at a time with the analysis being re-run with each item removal until a parsimonious scale structure was identified. Items with the greatest number of violations (e.g., those with loadings less than 0.40, multiple cross-loadings at 0.30 or greater, and poor theoretical fit) were removed first.[9] For analysis purposes, cross-loadings were defined as substantial (≥ 0.30) or extreme (≥ 0.45).[9,21] Lastly, Cronbach's alpha and McDonald's Omega were assessed on each factor and set a priori between 0.70 and 0.89.[9,10,13]

Scale Structure - Confirmatory Factor Analysis of the Alternate Model

The parsimonious solution identified during the EFA process was then assessed using CFA procedures in AMOS with the sample n2. The same goodness-of-fit criteria that were utilized for the initial CFA were also used to assess model fit.[12,21] Parallel analysis was also conducted as an additional method to determine the number of factors to retain.[27] Additionally, modification indices, factor loadings, and correlations between constructs were analyzed. Lastly, a correlational analysis was conducted on the scores of the HOOS and the alternate HOOS to determine if the scale explained and acceptable amount of variance ($r \geq 0.90$; $R^2 = 0.81$).[28,29]

Invariance Testing

Invariance testing was conducted if recommended model fit criteria were met to determine if the association between the latent constructs and the associated items were stable and equal across groups (i.e., age and sex) and time points.[12,21,30] This process requires using a set of hierarchical steps with an increasing levels of constraint.[12,21,30] Additionally, two types of invariance testing were conducted separately: 1) multi-group (i.e., age groups and sex) at time point 1 (i.e., preoperative visit), and 2) longitudinal across five time points (i.e., preoperatively [time point 1], 6 months postoperatively [time point 2], 1-year postoperatively [time point 3], 2-years postoperatively [time point 4], and 3-years postoperatively [time point 5]). First, the samples (i.e., n1 and n2) were combined back into one full sample. Individual CFAs were then conducted by subgroup category and across each time point, ensuring the construct and factors (e.g., pain, function, symptoms, etc.) were measuring what was intended among groups and across time.[21,30]

Following the individual CFAs, the model then underwent configural, metric, and scalar invariance testing.[20,21,30] First, the configural invariance test placed all groups and time points in the same model to ensure the same factors have similar items across groups (e.g., males and females) or time (e.g., time point 1 and time point 2). Secondly, the metric model then tested to determine if factor loadings were equal across groups and time points; invariance at this step would ensure that the meanings of the factors are equal across groups and time points.[20] If metric model invariance requirements were met, equal variances (i.e., group differences and time differences) were assessed.[20] Lastly, the scalar invariance test ensures that item intercepts were equal across groups and time points, which indicates the means were not determined or altered by external factors.[20] If scalar invariance requirements were met, equal mean models (i.e., score differences) were tested between groups and time points.[20]

The chi-square difference test (χ^2_{DIFF}) and the CFI difference test (CFI_{DIFF}) were used to compare model fit, with a p-value cut-off of 0.01.[25,30] Due to the χ^2_{DIFF} test being sensitive to sample size,[25] the CFI_{DIFF} test held greater weight in decisions regarding invariance testing model fit. If a model exceeded the χ^2_{DIFF} test, but met the CFI_{DIFF} test, invariance testing proceeded.

Latent Growth-Curve Modeling

If the model was deemed invariant across time, LGC modeling was conducted using Full Information Maximum Likelihood in AMOS.[12] Analysis using LGC models allows for the evaluation of model adequacy using model fit indices and model selection criteria, which accounts for measurement error using latent repeated-measures, and has the flexibility to deal effectively with missing data.[31,32] First, *a priori* was hypothesize to be linear,[20] meaning a steady decrease in pain and an increase function from preoperative to postoperative visits. The same goodness-of-fit criteria utilized for the CFAs were used to assess fit for the LGC model.[12,20,21,30]

The model included two growth parameters: a) the intercept parameter representing an individual's score preoperatively (time point 1), and b) the slope parameter representing the individual's rate of change over time.[20] Also, covariance between the intercept and slope factors was also included; this provides an indication of whether patients who started at a lower or higher score for the outcome changed at slower or higher rate.[20] The values assigned to the slope parameters were represented as the preoperative visit = 0, 6-months postoperatively = 0.5, 1-year postoperatively = 1, 2-year postoperatively = 2, and 3-year postoperatively = 3, respectively.[20,33] If the interindividual growth trajectory was statistically significant (i.e., the sample was heterogeneous), groups (i.e., age and sex) were then added to determine if the sample could be further explained.[12,20] If a nonlinear growth model was found, the slope parameters would be freely estimated to explore the shape of the growth model.[20,33]

Results

A total of 6724 complete responses (i.e., all items of the HOOS were answered preoperatively) were extracted for data cleaning. From the sample of 6724 cases, univariate (N = 148; 2.2%) and multivariate (N = 0; 0.0%) outliers were identified at the preoperative visit and were removed during the data cleaning process. Deleted cases included females (N = 68) and males (N = 80), with an average age of 67.36 ± 10.36y (range = 42-87). To perform the EFA, the data was then split into two random equal samples (n1 = 3288, n2 = 3288) of hip arthroplasty cases. The random selection of cases was generated in a three-step process: 1) a random number generator created a unique and random number for each arthroplasty case in the data set; 2) the unique identifiers were then sorted in ascending order; 3) the first 3288 cases were selected for n1, and the remainder were selected for n2. Sample n1 consisted of 48.1% males (n = 1582) and 51.7% females (n = 1701) with a mean age of 62.70 ± 11.17y (range = 23-89y). Sample n2 consisted of 48.1% males (n = 1582) and 51.7% females (n = 1701) with a mean age of 63.05 ± 10.69y (rang = 19-89y). Lastly, to perform the longitudinal invariance testing, only those individuals who responded to the 40-item HOOS at all five time points were included in the analysis. This sample consisted of 1144 patients (561 = males [49.0%], 580 = females [50.7%], 3 = missing [0.3%]) with a mean age of 62.03 ± 9.83y (range = 27-90y).

Scale Structure - Confirmatory Factor Analysis

The CFA conducted on the full sample (n = 6576) of the original five-factor 40-item HOOS did not meet recommended model fit indices values (CFI = 0.822, TLI = 0.809, IFI = 0.822, RMSEA = 0.085; Figure 5). Factor loadings were significant and ranged from 0.38-0.88. Additionally, correlations between first-order latent variables (e.g., symptoms, pain, ADLs) were high, ranging from r = 0.77-0.92. Modification indices revealed significant cross-loadings between several items (e.g., item 6 [how often is your hip painful] and item 37 [how often are you aware of your hip problem]; 536.30). Additionally, high error-term correlations were identified between several items (e.g., item 24 [putting on socks/stockings] and item 26 [taking off sock/stockings]; 2986.41). Therefore, the dataset was randomly split for further analyses due to poor model fit, the possible multicollinearity between the first-order latent variables, and item redundancy.

Alternate Model Generation

Initial EFA of the original HOOS in sample n1 extracted four factors with eigenvalues > 1.0 that accounted for 67.94% of the variance (Table 6). However, parallel analysis indicated three factors should be retained; the fourth factor was slightly under the recommended cut-off when the raw data was compared to the percentile random data eigenvalue (Table 7). Following extraction, item loadings, cross-loadings, and analysis of item content were assessed individually throughout the process; 25 items that had low loadings, substantial cross-loadings, or poor conceptual fit and were eliminated. An additional seven items were removed (i.e., 31 items in total) due to low loadings, high cross-loadings, inflated high inter-item correlation values, or lack of conceptual fit. When these items were removed, a three-factor, 9-item alternate HOOS (i.e., HOOS-9) was identified (Table 8); the solution accounted for 81.64% of the variance, contained items with loadings ≥ 0.48, and had Cronbach's alpha and MacDonald's Omega ranging from 0.86-0.89. Factor two and factor three accounted for more than 5% of the variance, but the eigenvalues fell below the *a priori* cut-off of 1.0.

Factor one contained the original HOOS items 17, 19 and 21 and assessed perceived function pertaining to daily living and were relabeled "ADLs." Factor two contained the original HOOS items 33, 34, and 35, which measured difficulty pertaining to higher levels of activity, and retained the original label "Sport." Lastly, factor three contained the original HOOS items 38, 39, and 40 that assess the patients' perceived hip related QOL; therefore, the original label "QOL" was retained.

Scale Structure - Confirmatory Factor Analysis of the Alternate Model

The alternate model of the HOOS-9 had an improved fit (Figure 6), with the goodness-of-fit indices exceeding the recommended values (CFI = 0.974, TLI = 0.961, RMSEA = 0.046), when sample n2 was used within the CFA procedures.[12,21] Factor loadings were significant and ranged from 0.49-0.75. Correlations between first-order latent variables (i.e., ADLs, Sport, QOL) were improved and ranged from r = 0.38-0.56. Modification indices did not reveal any significant cross-loadings or error-term correlations. Lastly, participant scores from original 40-item HOOS and the modified HOOS-9 were highly correlated between the function of daily living and ADLs constructs (r = 0.90, R^2 = 0.80), sport constructs (r = 0.97, R^2 = 0.94), and QOL constructs (r = 0.98, R^2 = 0.96).[30,31]

Multi-group Invariance Testing of the Alternate Model

Sex Groups

The complete sample was used to conduct invariance testing: of the 6566 individuals in the sample, 3164 (48.2%) were males and 3402 (51.8%) were females. The initial model (i.e., equal form) met all model fit indices (CFI = 0.978; χ^2 = 699.88; RMSEA = 0.064; Table 9). The metric model (i.e., equal loadings) passed both the CFI_{DIFF} test (CFI = 0.977) and the χ^2_{DIFF} test (χ^2 = 31.67). Because the metric model was invariant between groups, examination of the equal latent variable factors was warranted. The equal factor variance model passed the CFI_{DIFF} test (CFI = 0.977) and the χ^2_{DIFF} test (χ^2 = 42.67), indicating variances of the latent variables were equal between groups. The scalar model (i.e., equal intercepts) passed both the CFI_{DIFF} test (CFI = 0.975) and the χ^2_{DIFF} test (χ^2 = 116.03). Because the scalar model was invariant between groups, examination of the latent mean model was warranted. The equal latent means model also passed the CFI_{DIFF} test (CFI = 0.975) and the χ^2_{DIFF} test (χ^2 = 201.43) indicating there were no significant differences in latent means between groups.

Age Groups

Patients who had an age reported were included in the analysis and grouped by the following age ranges: < 45y (n = 374; 5.69%), 45-54y (n = 1,043; 15.86%), 55-64y (n = 2200; 33.45%), 65-74y (n = 1983; 30.16%), and ≥ 75y (n = 976; 14.84%). The initial model (i.e., equal form) met all model fit indices (CFI = 0.977; χ^2 = 801.94; RMSEA = 0.029; Table 10). The metric model (i.e., equal loadings) passed both the CFI_{DIFF} test (CFI = 0.977) and the χ^2_{DIFF} test (χ^2 = 41.39). Because the metric model was invariant between groups, examination of the equal latent variable factors was warranted. The equal factor variance model passed the CFI_{DIFF} test (CFI = 0.975) and the χ^2_{DIFF} test (χ^2 = 99.43), indicating variances of the latent variables were equal between groups. The scalar model (i.e., equal intercepts) passed both the CFI_{DIFF} test (CFI = 0.971) and the χ^2_{DIFF} test (χ^2 = 248.67). Because the scalar model was invariant between groups, examination of the latent mean model was warranted. The equal latent means model also passed the CFI_{DIFF} test (CFI = 0.969) and the χ^2_{DIFF} test (χ^2 = 301.23) indicating there were no significant differences in means between groups.

Longitudinal Invariance Testing of the Alternate Model

Of the total 6576 patients included in the full sample, 1144 (17.4%) patients responded to the HOOS at all five time points (i.e., preoperatively, 6-months postoperatively, 1-year postoperatively, 2-years postoperatively, 3-years postoperatively). The initial model (i.e., equal form) met all model fit indices (CFI = 0.979; χ^2 = 1511.52; RMSEA = 0.030; Table 11). The metric model (i.e., equal loadings) passed both the CFI_{DIFF} test (CFI = 0.979) and the χ^2_{DIFF} test (χ^2 = 57.53). Because the metric model was invariant between groups, examination of the equal latent variable factors was warranted. The equal factor variance model passed both the CFI_{DIFF} test (CFI = 0.973) and the χ^2_{DIFF} test (χ^2 = 262.11), indicating variances of the latent variables were equal between groups. The scalar model (i.e., equal intercepts) also met the CFI_{DIFF} test (CFI = 0.973) and the χ^2_{DIFF} test (χ^2 = 299.08). Because the scalar model passed the CFI_{DIFF} and the χ^2_{DIFF} test, evaluation of the equal latent means model was warranted. The equal latent means model did not pass the CFI_{DIFF} test (CFI = 0.913) or the χ^2_{DIFF} test (χ^2 = 2504.84), indicating that the means between time points were significantly different. When the means were not constrained to be equal, the mean scores significantly improved over each time point (i.e., less hip disability) with time 5 having an overall lower score compared to baseline.

Latent Growth-Curve Model of the Alternate Model

Patients who answered the HOOS at all five time points (N = 1144) were included in the analyses. In addition, the scores of the alternate HOOS-9 were calculated by averaging the nine items, dividing by 4, multiplying by 100, then subtracting the total from 100. These calculation guidelines are similar to the guidelines published by previous authors[34]; however, a total score was calculated for the alternate HOOS-9 rather than individual construct scores (i.e., ADLs, Sport, QOL). In addition, a bifactor model was assessed to determine if the scale could be summed as a total score rather than the individual construct scores.[20] Goodness-of-fit indices met the recommended criterion (CFI = 0.995, TLI = 0.988, IFI = 0.995, RMSEA = 0.026).[20] Therefore, we proceeded with the total summed score for assessment of LGC modeling.

The linear LGC model did not meet the recommended model fit indices (CFI = 0.110, TLI = 0.110, IFI = 0.110, RMSEA = 0.441; Figure 7).[20] These findings suggest that patients' response scores increase over each time point (i.e., hip disability improving over time), however the change in scores were not increasing at a consistent rate over each visit. Therefore, exploratory methods were used to determine the shape of the growth trajectory. To perform this, slope parameters for time points 1 and 2 were constrained to 0 and 0.50, respectively. However the remainder of the time points (i.e., time points 3-5) were freely estimated to explore the shape of the growth model.[20,33] Upon assessment of these findings, the final slope parameters were defined as follows: preoperatively = 0, 6-months postoperatively = 0.50, 1-year postoperatively = 0.52, 2-years postoperatively = 0.54, and 3-years postoperatively = 0.55. When the parameters were constrained to the prior definitions, the model met most recommended model fit indices (CFI = 0.970, TLI = 0.970, IFI = 0.970; Figure 8); however, the RMSEA was slightly exceeded (0.080). Upon assessment of the means estimates, the parameters for both the intercepts and the shape were statistically significant ($p < 0.001$). In addition, findings revealed the average score for the alternate HOOS-9 to be 37.55 points at time point 1 and that scores increased over the 3-year period to 87.33 points. When assessing the covariance between the intercept and shape, a negative estimate was identified (i.e., -77.42). Upon assessment of the model estimates, the variances were not statistically significant for intercept ($p = 0.132$) or shape ($p = 0.061$).

Model fit indices were also assessed between groups (i.e., sex and age) to determine the differences sex and age groups have on the mean scores and growth trajectories over time. When assessing the LGC model pertaining to sex, the model met all recommended fit indices (CFI = 0.970, TLI = 0.970, IFI = 0.970, RMSEA = 0.057). Estimates pertaining to the means for both the intercept and the shape were statistically significant ($p < 0.001$). Upon assessment of the mean scores, the average mean score for the alternate HOOS-9 at preoperative visit were higher in patients within the male group (40.63) compared to patients in the female group (34.56). In addition, patients in both groups improved their scores over the 3-years, with patients in the female group having an overall higher change in scores (i.e., females = 92.86, males = 91.75). When assessing the covariance between the intercept and shape, a negative estimate was identified for patients in the male group (i.e., -67.61) and female group (-52.41; Table 12). When assessing the variances of the model, estimates were not statistically significant for the intercept ($p = 0.282$) or shape ($p = 0.202$).

When assessing the LGC model for age groups, the model met all recommended fit indices (CFI = 0.973, TLI = 0.977, IFI = 0.973, RMSEA = 0.035). Estimates pertaining to the means for both the intercept and the shape were statistically significant ($p < 0.001$). On average, patients in the age group 45-54 had an overall lower score at the preoperative visit (33.42) when compared to patients in the other age groups (Table 12). In addition, patients in all groups increased their scores over the 3-years, with patients in the 45-54 group having an overall higher change in scores (91.90). When assessing the covariance between the intercept and shape, those aged < 45 had a large negative estimate (-461.30) compared to the other age groups (Table 12). Pertaining to the variances of the model, estimates were not significantly different for the intercept ($p = 0.336$) or shape ($p = 0.187$).

Discussion

With the occurrence of THAs expected to significantly increase by 2050,[35] it is imperative that clinicians have access to PROs that can be widely used across different sexes (i.e., males and females), age groups (i.e., 18-94), and repeated visits. Having a PRO to assess the patient perspective of hip health throughout a patients' recovery is beneficial to clinicians to ensure positive outcomes following arthroplasty. More recently, significant attention has been focused on PROs associated with outcomes following THA.[36,37] Therefore, the need for establishing a psychometrically sound tool to adequately measure the multifaceted nature of hip pain and function is valuable. Previous psychometric analysis on the HOOS has not yielded a scale structure that meets recommended model fit indices,[16] and assessment of how age and sex influence patient responses to the individual items and mean scores has not been condcuted.[1,16,38] As such, the primary purpose of this study was to assess the psychometric properties of the HOOS in patients undergoing a THA. The CFA of the original 40-item HOOS did not meet recommended model fit indices. Therefore, an EFA was conducted to establish a more parsimonious scale structure. The alternate three-factor, 9-item (HOOS-9) was then subjected to multi-group invariance testing (i.e., sex and age groups), longitudinal invariance testing over five time points (i.e., preoperatively, 6-months postoperatively, 1-year postoperatively, 2-years postoperatively, 3-years postoperatively), and LGM over the five time-points. The alternate HOOS-9 met recommended measurement criteria, and additionally, can be recommended for use in research and clinical practice.

Scale Structure - Confirmatory Factor Analysis

The original five-factor, 40-item scale structure was not supported in our study due to poor model fit indices and high latent variable correlations. Our findings are consistent with our previous research where a well-supported scale structure in mostly healthy adults was not found.[16] High to very high correlations ($r = 0.77$-0.91) between the first-order latent variables were found, which indicates the potential of multicollinearity between the factors. Modification indices also suggested there were items with meaningful cross-loadings, indicating overlapping items were present (e.g., item six [how often is your hip painful] and item 37 [how often are you aware of your hip problem]) and high error-term correlations between several items (e.g., item 24 [putting on socks/stockings] and item 26 [taking off sock/stockings]). These findings further suggest the presence of multicollinearity; poor model fit, and the presence of multicollinearity, provide evidence that the scale should not be used in its current form.[12,21,39] Thus, scale refinement using alternate model generation was warranted to determine whether a psychometrically sound version could be identified using the current items.[12,21,39]

Psychometric Analysis of the Alternate HOOS-9

An EFA was conducted in a calibration sample (i.e., n1) and an alternate three-factor, 9-item solution (i.e., HOOS-9) was identified. The nine items represented three of the original five constructs of the HOOS: three items from “Function, daily living,” three items from “Function, sports and recreational activities,” and three items from “Quality of Life.” The alternate HOOS-9 was then subjected to covariance modeling procedures using the validation sample (i.e., n2). Though the alternate HOOS-9 only retained 22.5% of the questions from the original scale, participant responses were highly correlated ($r = 0.934$) with the original 40-item HOOS and accounted for a substantial amount of the variance ($R^2 = 0.872$).

The three-factor structure identified in our sample is different than other HOOS short-forms including the HOOS-JR, HOOS-PS, and more specifically the three-factor, 12-item HOOS (HOOS-12).[15,40] The HOOS-12 is short-form version of the original 40-item HOOS that includes three-factors (i.e., Pain, Function daily living, and QOL) which consist of 12 items (i.e., original HOOS items 6, 9, 10,12, 18, 19, 22, 36-40); however, our alternate HOOS-9 model contains four items present in the HOOS-12 (i.e., 19, 38, 39, 40) in the ADLs and QOL construct. Upon development of the HOOS-12, authors using computerized adaptive test (CAT) simulations to identify items to best match patients’ level of pain and function.[40] Limitations exist with the use of CAT such as high cost and the adaptability of the questionnaire to the individual persons responses.[41] Therefore, patients may not be answering the same questions based on their responses to the bank of items. This methodology poses further limitations on the ability of clinicians attempting to draw conclusions of the PROs; as such, CAT may not always be appropriate in the clinical setting.[41] Additional assessment of structural validity on the HOOS-12 was conducted by performing CFAs on the individual constructs (i.e., pain, function, QOL) and not on the full scale.[15] Best practice recommendations when performing CFA is to assess the entire scale and if the model meets recommended fit indices, perform invariance testing (e.g., multigroup, longitudinal) to ensure the instrument can be used across several groups and time.[12,21,30]

In addition to these findings, previous research assessing structural validity using CFA on the full HOOS-12 did not support its use in a sample of mostly healthy adults.[16] Several concerns related to the scale were noted: high correlations between the constructs (i.e., indicating potential multicollinearity), high Cronbach's alpha values (i.e., indicating potential item redundancy), cross-loadings of items (i.e., items shared commonalities).[16,21] Therefore, further testing of the HOOS-12 was not warranted in the population studied.[16] In our identified model, correlations between constructs ranged from 0.38-0.56, and modification indices did not reveal cross-loadings between items. Therefore, our findings present a newly refined short-form version of original HOOS items that measures unique constructs.

Multi-group and Longitudinal Invariance Testing of the Alternate HOOS-9

We assessed group differences using CFA methods between groups of interest (i.e., age groups and sex) and across several time points for the alternate HOOS-9. Invariance testing confirms the structural validity of the scale, ensuring the association between constructs (i.e., ADLs, Sport, and QOL) are being measured and their items are being interpreted similarity across groups (i.e., males, females) and time (i.e., multiple visits).[12,20,21] Thus, an invariant instrument allows clinicians to compare scores across groups or visits, and provides support that score differences in hip health are true group differences as opposed to measurement error.[12,21] Minimal studies exist assessing multi-group and longitudinal invariance testing using any forms of the HOOS; our previous work has focused on invariance testing pertaining to multiple short-forms (i.e., the HOOS-JR and HOOS-PS).[16] In a previous study, however, we assessed differences between hip pathology and physical activity groups in the HOOS-JR and HOOS-PS.[16] In a more recent study, we assessed multi-group (i.e., age groups and sex) and longitudinal invariance (i.e., multiple visits) in a similar sample of patients who underwent a THA. To our knowledge, this was the second study to assess multi-group and longitudinal invariance in a short-form version (i.e., HOOS-9) of the original 40-item HOOS.

We found the alternate HOOS-9 was invariant at the preoperative visit (i.e., preoperative THA) between age groups (i.e., < 45, 45-54, 55-64, 65-74, ≥ 75) and sex (i.e., males, females). These results indicate that the newly modified scale can be used to assess differences in hip-related dysfunction in patients undergoing a THA. In addition to our invariant findings, significant latent variances and latent mean differences were not found between age groups or sex suggesting minimal differences in hip disability were perceived between groups. These findings are different than our previous research where we identified latent mean differences in sex groups with females reporting higher mean scores on the HOOS-JR compared to males. In addition, other researchers identified sex and age differences, with females and those in older age groups reporting higher scores on the 40-

item HOOS and HOOS-12 for all domains.[42,43] However, Sunden et al. only identified significant differences between males and females in the oldest age group (i.e., 75-84); no significant differences in mean scores were found between males and females in different age groups (i.e., 18-35, 3-54, 55-74).[42] Larsen et al. found significant worse HOOS and HOOS-12 scores with increasing age. Within our population, the majority of our sample was younger than 75 years of age which could partially explain these findings. Of important note, these findings, however, are associated with different versions of the HOOS scales which include different items. Having different items compared to the other versions indicates the scales are not necessarily measuring hip disability in the exact same way. Therefore, our findings are unique in that the scale structure of the alternate HOOS-9 demonstrates no significant differences between sex and age groups.

This study also provides evidence of scale validity of the alternate HOOS-9 for assessing postoperative effects across time. Longitudinal invariance was established across multiple visits (i.e., preoperatively and 6-months, 1-year, 2-years, and 3-years postoperatively), indicating the scale can be used to assess differences in hip disability across multiple visits. Thus, the results supported the assessment of mean scores across time to determine if scores change post-THA. We identified significant latent mean differences were identified across time points, indicating patients reported a meaningful improvement in scores from preoperatively to 3-years postoperatively. In addition, the highest scores (i.e., more hip disability) were reported preoperatively and the lowest scores (i.e., less hip disability) were identified at 3-years postoperatively. These findings provide support for scale validity as patients who receive surgery would be expected to report improvement over time following the intervention (i.e., THA) as natural healing occurs across visits. These findings are congruent with previous research reporting significant improvement in scores on the HOOS and HOOS-12 in patients who underwent a THA from preoperatively to 2-years postoperatively.[44,45]

Alternate HOOS-9 Latent Growth-Curve Modeling

To our knowledge, this was the first study to perform LGC modeling in patients who answered questions of the 40-item HOOS over a 3-year period postoperatively. Use of LGC modeling is a robust technique that allows researchers to assess between-person differences in within-person change which is unique compared to traditional longitudinal assessments (e.g., repeated-measures analyses or multivariate analyses).[20,46] In addition, LGC modeling is highly flexible when attempting to assess differences in unequally spaced time points (e.g., months, years) and for more complex nonlinear data.[20,46] Few studies identified assessed outcomes related to hip disability (i.e., HOOS Physical Function [HOOS-PS], Oxford Hip Score [OHS]) over a 12-month period postoperatively in patients who underwent a THA.[47,48] In addition, other researchers assessed LGC of the OHS in patients over 6-weeks postoperatively.[49] These three studies all identified a nonlinear improvement; most improvement occurred within the first 6-weeks to 3-months postoperatively.[47-49] These findings are similar to ours; the lack of fit within the linear model, along with the re-defined nonlinear model, demonstrates the majority of the growth and improvement in scores occurred within the first 6-months postoperatively.

Researchers defined groups by healing trajectories (e.g., fast starters, early recovery),[48,49] PROs (i.e., OHS, HOOS-PS), or how the patients scored (i.e., high-high, intermediate, low-high).[47] These defined groups differ from our study examining the differences age groups and sex have on responses to the alternate HOOS-9 over time. Our results indicate that patients in the male group have an overall higher score at baseline (40.63) compared to those in the female group (34.56). In addition, patients in both male and female groups who scored lower at baseline had an overall faster growth over time, although those in the female group had a slower rate of growth over time in comparison to males (-52.41 vs. -67.61), respectively. These findings are similar to Hesseling et al. who demonstrated that females were considered a slow starter, meaning they had slower improvement in hip function and QOL within the first 3-months postoperatively but had an overall improvement at 1-year postoperatively.[48] In addition, we found females had a higher mean score at 3-years postoperatively (92.86) in comparison to males (81.75). However, even though differences were identified between patients in the male and female groups, the variances of the model for the intercept and shape were not statistically significant. This finding indicates that there were no significant differences between the two groups (i.e., interindividual differences) noted.

To our knowledge this was also the first study to assess different age groups across the time points. When assessing these differences, patients in the age group 45-54 scored the lowest overall at baseline (33.42) when compared to the other groups, and patients in the age group ≥ 75 had the highest overall mean score at baseline (41.12). In addition, patients in the age group > 45 who had lower self-perceived hip function and QOL made greater improvements in their scores (-461.30) compared than those in the aged 45-54 (-103.42) and 65-74 (-342.90); however, they had a slower rate of increase in scores over time. Patients in the ≥ 75 group had a steeper growth and improvement in HOOS-9 scores (451.13) over time, though had an overall lower mean score at time 5 (82.67) in comparison to those in the age group 55-64 (133.96). These findings indicate that patients in the age group ≥ 75 improve their hip disability and QOL faster but have an overall lower score on the alternate HOOS-9 compared to the other age groups. Variances between the intercept and slope, however, were not statistically significant ($p > 0.05$), which indicate interindividual differences are homogenous rather than heterogeneous.

Limitations and Future Research

Although our study included a large sample of patients undergoing a THA, there are limitations that should be addressed. Even though the alternate HOOS-9 was assessed using a cross-validation sample with the decision to split the sample, the participants used had responded to the original 40-item HOOS. As such, the responses to the alternate HOOS-9 items could have been influenced by the other 31 items.[39] Therefore, future research should be conducted on a sample of individuals who only respond to the nine items.[39] We assessed concurrent validity (i.e., correlation between two scales) between the original constructs of the 40-item HOOS and the newly proposed scale. Future researchers may want to consider conducting further analyses that correlate the HOOS-9 responses with other scales designed to measure similar dimensions (e.g., QOL). As this is the first study to report the HOOS-9, limitations may exist when attempting to assess differences in clinical practice and research. Therefore, future research should be conducted to determine the responsiveness, minimal clinically important difference and reliability of the instrument.

Additionally, even though we found invariance between groups of interest (i.e., sex and age groups), we were unable to capture other pertinent information such as demographic data (e.g., race, ethnicity, medical history), diagnosis (e.g., osteoarthritis, hip dysplasia), surgical procedure (i.e., primary, revision), or operative data (e.g., surgical approach, laterality, implant type) due to a high amount of missing or uncollected data from the SOS database. Thus, caution is warranted when examining alternate HOOS-9 differences in groups that have not yet been analyzed. Future research should focus on invariance testing modeling across several different groups (e.g., diagnosis, surgical approach, surgical procedure) to ensure the scale has the necessary properties to support between groups analysis in these populations. In addition, further analyses using LGC modeling within these different groups could help clinicians and researchers understand healing differences over time.

Another limitation of this study was the decision to score the alternate HOOS-9 as a total score versus scoring each construct individually for purposes of LGC modeling. Scoring PROs as a total score is common practice for documentation purposes to be able to easily assess changes over time. As our model fit statistics reveal low to moderate correlations between the first-order latent variables, this provides justification that the items are measuring unique constructs. However, we performed a bifactor model to determine if a composite score could be used even though the constructs were unique. Our findings reveal acceptable goodness-of-fit indices indicating clinicians may be able to score the alternate HOOS-9 as a total summed score. Therefore, future research should be conducted to assess the reliability and validity (e.g., responsiveness) of the alternate HOOS-9 using the total summed scores. Although we had an overall large sample for this study, the sample size was much smaller (n = 1140) when assessing invariance and differences over time (i.e., longitudinal invariance and LCG modeling) due to the low percentage (17.4%) of patients who answered the items over all time points. Therefore, future research should be conducted in a larger sample to ensure similar findings exist.

Conclusion

The original scale structure of the 40-item HOOS was not supported in our study. We subsequently identified an alternate three-factor, 9-item HOOS (i.e., HOOS-9) that met contemporary model fit recommendations, along with multi-group and longitudinal invariance testing. Our findings support the use of the alternate HOOS-9 as a more viable option to assess hip disability in research and clinical practice, but caution is warranted until more research is conducted to further assess the measurement properties of the alternate HOOS-9 scale.

References

1. Nilsdotter AK, Lohmander LS, Klässbo M, Roos EM. Hip disability and osteoarthritis outcome score (HOOS)–validity and responsiveness in total hip replacement. *BMC Musculoskelet Disord.* 2003;4(1):10.

2. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. 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.* 1988;15(12):1833-1840.

3. Roos EM, Roos HP, Lohmander LS, Ekdahl C, Beynnon BD. Knee Injury and Osteoarthritis Outcome Score (KOOS)—development of a self-administered outcome measure. *J Orthop Sports Phys Ther.* 1998;28(2):88-96.

4. A User's Guide to: Hip disability and Osteoarthritis Outcome Score HOOS http://www.koos.nu/. Published 2003. Updated 2008. Accessed July 25, 2021.

5. Centers for Medicare and Medicaid Services. Measure Methodology. Osteoarthritis (OA): Function and Pain Assessment Web site. https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Measure-Methodology. Published 2021. Accessed July 25, 2021.

6. De Groot I, Reijman M, Terwee C, et al. Validation of the Dutch version of the Hip disability and Osteoarthritis Outcome Score. *Osteoarthritis Cartilage.* 2007;15(1):104-109.

7. 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.

8. Ornetti P, Parratte S, Gossec L, et al. Cross-cultural adaptation and validation of the French version of the Hip disability and Osteoarthritis Outcome Score (HOOS) in hip osteoarthritis patients. *Osteoarthritis Cartilage.* 2010;18(4):522-529.

9. Leech NL, Barrett KC, Morgan GA. *IBM SPSS for intermediate statistics: Use and interpretation.* Routledge; 2014.

10. Taber KS. The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education.* 2018;48(6):1273-1296.

11. Pesudovs K, Burr JM, Harley C, Elliott DB. The development, assessment, and selection of questionnaires. *Optom Vis Sci.* 2007;84(8):663-674.

12. Kline RB. *Principles and practice of structural equation modeling.* Guilford publications; 2015.

13. Streiner DL. Starting at the beginning: an introduction to coefficient alpha and internal consistency. *J Pers Assess.* 2003;80(1):99-103.

14. Panayides P. Coefficient alpha: interpret with caution. *Eur J Psychol.* 2013;9(4):687–696-687–696.

15. Gandek B, Roos E, Franklin PD, Ware Jr JE. Item selection for 12-Item short forms of the knee injury and osteoarthritis outcome score (KOOS-12) and hip disability and osteoarthritis outcome score (HOOS-12). *Osteoarthritis Cartilage.* 2019;27(5):746-753.

16. Miley EN, Casanova MP, Cheatham SW, Larkins L, Pickering MA, Baker RT. Confirmatory Factor Analysis of the Hip Disability and Osteoarthritis Outcome Score (HOOS) and Associated Sub-scales. *Int J Sports Phys Ther.* 2023.

17. Bilbao A, Quintana JM, Escobar A, Las Hayas C, Orive M. Validation of a proposed WOMAC short form for patients with hip osteoarthritis. *Health Qual Life Outcomes.* 2011;9(1):1-11.

18. Stratford P, Kennedy D, Woodhouse L, Spadoni G. Measurement properties of the WOMAC LK 3.1 pain scale. *Osteoarthritis Cartilage.* 2007;15(3):266-272.

19. Gerbing DW, Anderson JC. On the meaning of within-factor correlated measurement errors. *J Consum Res.* 1984;11(1):572-580.

20. Byrne BM. *Structural equation modeling with AMOS: basic concepts, applications, and programming (multivariate applications series)*. Taylor & Francis Group; 2010.

21. Brown TA. *Confirmatory factor analysis for applied research.* Guilford publications; 2015.

22. Bryant FB, Yarnold PR. *Principal-components analysis and exploratory and confirmatory factor analysis*. 1995.

23. Hayton JC, Allen DG, Scarpello V. Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organ Res Methods.* 2004;7(2):191-205.

24. Tabachnick BG, Fidell LS. *Using multivariate statistics.*Allyn & Bacon; 2001.

25. Grimm LG, Yarnold PR. *Reading and understanding multivariate statistics.* American Psychological Association; 1995.

26. Schönrock-Adema J, Heijne-Penninga M, Van Hell EA, Cohen-Schotanus J. Necessary steps in factor analysis: enhancing validation studies of educational instruments. *Med Teach.* 2009;31(6):e226-e232.

27. O'connor BP. SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behav Res Methods Instrum Comput.* 2000;32(3):396-402.

28. Ware Jr JE, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Medical Care.* 1996:220-233.

29. Raes F, Pommier E, Neff KD, Van Gucht D. Construction and factorial validation of a short form of the self-compassion scale. *Clin Psychol Psychother.* 2011;18(3):250-255.

30. Byrne BM. *Structural equation modeling with Mplus: Basic concepts, applications, and programming.* Routledge; 2013.

31. Preacher KJ, Wichman AL, MacCallum RC, Briggs NE. *Latent growth curve modeling.* Sage; 2008.

32. Grimm KJ, Ram N, Estabrook R. *Growth Modeling: Structural Equation and Multilevel Modeling Approaches.* Guilford Publications; 2016.

33. Burant CJ. Latent growth curve models: Tracking changes over time. *Int J Aging Hum Dev.* 2016;82(4):336-350.

34. Hip diability and Osteoarthritis Outcome Score (HOOS) Scoring Instructions. Mapi Research Trust. Published 2013. Accessed June 3, 2023.

35. Centers for Disease Control and Prevention. A National Public Health Agenda for Osteoarthritis: 2020 Update. https://www.cdc.gov/arthritis/docs/oaagenda2020.pdf. Published 2020. Accessed November 6, 2020.

36. Centers for Medicare and Medicaide Services. FY 2023 Hospital Inpatient Prospective Payment System (IPPS) and Long-Term Care Hospital Prospective Payment System (LTCH PPS) Final Rule — CMS-1771-F. https://www.cms.gov/newsroom/fact-sheets/fy-2023-hospital-inpatient-prospective-payment-system-ipps-and-long-term-care-hospital-prospective. Published 2022. Accessed December 31, 2022.

37. Comprehensive Care for Joint Replacement Model. CMS.gov. https://innovation.cms.gov/innovation-models/cjr. Accessed 11.27.22.

38. Arbab D, van Ochten JH, Schnurr C, Bouillon B, König D. Assessment of reliability, validity, responsiveness and minimally important change of the German hip dysfunction and osteoarthritis outcome score (HOOS) in patients with osteoarthritis of the hip. *Rheumatol Int.* 2017;37(12):2005-2011.

39. Dillman DA, Smyth JD, Christian LM. *Internet, phone, mail, and mixed-mode surveys: the tailored design method.* John Wiley & Sons; 2014.

40. Gandek B, Roos EM, Franklin PD, Ware Jr JE. A 12-item short form of the Hip disability and Osteoarthritis Outcome Score (HOOS-12): tests of reliability, validity and responsiveness. *Osteoarthritis Cartilage.* 2019;27(5):754-761.

41. Cook KF, O'Malley KJ, Roddey TS. Dynamic assessment of health outcomes: time to let the CAT out of the bag? *Health Services Research.* 2005;40(5p2):1694-1711.

42. Sundén A, Lidengren K, Roos E, Lohmander L, Ekvall Hansson E. Hip complaints differ across age and sex: a population-based reference data for the Hip disability and Osteoarthritis Outcome Score (HOOS). *Health Qual Life Outcomes.* 2018;16:1-6.

43. Larsen P, Rathleff MS, Roos EM, Elsoe R. National population-based reference data for the Hip Disability and Osteoarthritis Outcome Score (HOOS). *Arch Orthop Trauma Surg.* 2023:1-10.

44. Lyman S, Lee Y-Y, McLawhorn AS, Islam W, MacLean CH. What are the minimal and substantial improvements in the HOOS and KOOS and JR versions after total joint replacement? *Clin Orthop Relat Res.* 2018;476(12):2432.

45. Goodman SM, Mehta BY, Mandl LA, et al. Validation of the hip disability and osteoarthritis outcome score and knee injury and osteoarthritis outcome score pain and function subscales for use in total hip replacement and total knee replacement clinical trials. *J Arthroplasty.* 2020;35(5):1200-1207.

46. Curran PJ, Obeidat K, Losardo D. Twelve frequently asked questions about growth curve modeling. *Cognitive Development Society.* 2010;11(2):121-136.

47. Golinelli D, Grassi A, Tedesco D, et al. Patient reported outcomes measures (PROMs) trajectories after elective hip arthroplasty: a latent class and growth mixture analysis. *J Patient Rep Outcomes.* 2022;6(1):95.

48. Hesseling B, Mathijssen N, Van Steenbergen L, Melles M, Vehmeijer S, Porsius J. Fast starters, slow starters, and late dippers: trajectories of patient-reported outcomes after total hip arthroplasty: results from a Dutch nationwide database. *J Bone Joint Surg.* 2019;101(24):2175-2186.

49. Porsius JT, Mathijssen NM, Klapwijk-Van Heijningen LC, Van Egmond JC, Melles M, Vehmeijer SB. Early recovery trajectories after fast-track primary total hip arthroplasty: the role of patient characteristics. *Acta Orthop.* 2018;89(6):597-602.

Tables

Table 6

Initial EFA Extraction of the 40-item HOOS in Sample n1

Item	Factor 1	Factor 2	Factor 3	Factor 4
Pre_HOOSQ21	**0.921**			
Pre_HOOSQ14	**0.904**			
Pre_HOOSQ9	**0.872**			
Pre_HOOSQ19	**0.740**			
Pre_HOOSQ15	**0.670**			
Pre_HOOSQ23	**0.667**			
Pre_HOOSQ12	**0.645**			
Pre_HOOSQ16	**0.594**			
Pre_HOOSQ10	**0.570**			
Pre_HOOSQ17	**0.562**			
Pre_HOOSQ32	**0.561**			
Pre_HOOSQ36	**0.522**			0.351
Pre_HOOSQ3	**0.436**			0.356
Pre_HOOSQ26		**0.778**		
Pre_HOOSQ24		**0.772**		
Pre_HOOSQ30		**0.524**		
Pre_HOOSQ20		**0.500**		
Pre_HOOSQ22		**0.477**		
Pre_HOOSQ25	0.313	**0.466**		
Pre_HOOSQ28	0.326	**0.420**		
Pre_HOOSQ18		**0.377**		
Pre_HOOSQ13			**-0.837**	
Pre_HOOSQ11			**-0.808**	
Pre_HOOSQ29			**-0.518**	
Pre_HOOSQ27		0.336	**-0.513**	
Pre_HOOSQ5			**-0.472**	
Pre_HOOSQ7	0.36		**-0.399**	
Pre_HOOSQ6			**-0.364**	
Pre_HOOSQ4			**-0.361**	

(table 6 continues)

Pre_HOOSQ2			**-0.355**	0.35
Pre_HOOSQ8			**-0.314**	
Pre_HOOSQ34				**0.659**
Pre_HOOSQ35				**0.597**
Pre_HOOSQ33		0.31		**0.547**
Pre_HOOSQ38				**0.518**
Pre_HOOSQ39				**0.451**
Pre_HOOSQ40				**0.449**
Pre_HOOSQ37				**0.403**
Pre_HOOSQ31	0.304			**0.401**
Eigenvalue	24.57	1.42	1.18	1.06
% Variance	61.43	3.56	2.96	2.66

Extraction Method: Maximum Likelihood; Rotation: Oblimin with Kaiser Normalization; **Bolded** values show the item loading for each factor.

Table 7

Parallel Analysis of Raw Data Eigenvalues, Means, and Percentile Random Data Eigenvalues

Number of items	Raw Data	Means	Random Data
1	24.57*	1.21	1.23*
2	1.42*	1.19	1.21*
3	1.18*	1.17	1.18*
4	1.06	1.16	1.17
5	0.79	1.15	1.16
6	0.75	1.13	1.15
7	0.71	1.12	1.13
8	0.63	1.11	1.12
9	0.58	1.10	1.11
10	0.51	1.09	1.10

* $p < 0.05$; Note: Table only presents data for the first 10 of 44 items

Table 8

Alternate HOOS-9 Exploratory Factor Analysis

Item	Factor 1	Factor 2	Factor 3
Pre_HOOSQ17	0.747		
Pre_HOOSQ19	0.944		
Pre_HOOSQ21	0.889		
Pre_HOOSQ33		0.762	
Pre_HOOSQ34		0.988	
Pre_HOOSQ35		0.747	
Pre_HOOSQ38			0.970
Pre_HOOSQ39			0.843
Pre_HOOSQ40			0.575
Eigenvalue	5.84	0.84	0.67
% Variance	64.89	9.29	7.46
Cronbach Alpha	0.89	0.86	0.88
MacDonald's Omega	0.89	0.86	0.88

Extraction Method: Maximum Likelihood; Rotation Method: Oblimin with Kaiser Normalization.

Table 9

Goodness-of-Fit Indices for Multi-Group Invariance Across Sex

Modified 9-item HOOS	χ^2	**df**	χ^2_{diff} (df_{diff})	**CFI**	CFI_{diff}	**TLI**	**RMSEA**
Males (n = 3164)	332.98	24	----	0.979	----	0.969	0.064
Females (n = 3402)	366.90	24	----	0.977	----	0.966	0.065
Configural (equal form)	699.88	48	----	0.978	----	0.967	0.045
Metric (equal loadings)	731.55	54	31.67 (6)	0.977	0.001	0.970	0.044
Equal factor variances*	742.55	57	42.67 (9)	0.977	0.001	0.971	0.043
Scalar (equal indicator intercepts)	815.91	60	116.03 (12)	0.975	0.003	0.970	0.044
Equal latent means*	853.31	63	201.43 (15)	0.975	0.006	0.972	0.044

* = Substantive questions

Table 10

Goodness-of-Fit Indices for Multi-Group Invariance Across Age

Modified 9-item HOOS	χ^2	**df**	χ^2_{diff} (df_{diff})	**CFI**	CFI_{diff}	**TLI**	**RMSEA**
< 45 (n = 374)	77.48	24	----	0.966	----	0.950	0.080
45-54 (n = 1043)	167.60	24	----	0.968	----	0.952	0.076
55-64 (n = 2200)	232.38	24	----	0.980	----	0.971	0.063
65-74 (n = 1983)	197.52	24	----	0.980	----	0.971	0.060
≥ 75 (n = 976)	126.86	24	----	0.977	----	0.965	0.066
Configural (equal form)	801.94	120	----	0.977	----	0.966	0.029
Metric (equal loadings)	843.33	144	41.39 (24)	0.977	0.000	0.971	0.027
Equal factor variances*	901.37	156	99.43 (36)	0.975	0.002	0.971	0.027
Scalar (equal indicator intercepts)	1050.61	177	248.67 (57)	0.971	0.006	0.974	0.026
Equal latent means*	1103.17	180	301.23 (60)	0.969	0.008	0.969	0.028

* = Substantive questions

Table 11

Goodness-of-Fit Indices for Longitudinal Invariance Across Time Points

Modified 9-item HOOS	χ^2	***df***	χ^2_{diff} (df_{diff})	**CFI**	$\mathbf{CFI_{diff}}$	**TLI**	**RMSEA**
Preoperative	119.60	24	----	0.984	----	0.976	0.059
6-month postoperative	128.83	24	----	0.981	----	0.971	0.062
1-year postoperative	139.98	24	----	0.981	----	0.971	0.065
2-year postoperative	138.27	24	----	0.982	----	0.973	0.065
3-year postoperative	124.75	24	----	0.985	----	0.977	0.061
Configural (equal form)	1511.52	750	----	0.979	----	0.973	0.030
Metric (equal loadings)	1560.86	774	57.53 (24)	0.979	0.000	0.973	0.030
Equal factor variances*	1773.63	786	262.11(36)	0.973	0.006	0.966	0.033
Scalar (equal indicator intercepts)	1810.65	798	299.08 (48)	0.973	0.006	0.966	0.033
Equal latent means*	4016.36	810	**2504.84 (60)**	0.913	**0.066**	0.894	0.059

***** = Substantive questions; **Bolded** = did not meet cuff off criteria.

Table 12

Goodness-of-Fit Indices for Multi-Group Latent Growth-Curve Model

Modified 9-item HOOS	χ^2	*df*	CFI	TLI	RMSEA	Intercept Mean	Covariance
Males (n = 561)	40.89	10	0.976	0.976	0.074	40.63	-67.61
Females (n = 580)	53.35	10	0.963	0.963	0.087	34.56	-52.41
>45 (n = 48)	12.17	10	0.989	0.989	0.068	34.44	-461.30
45-54 (n = 205)	15.16	10	0.990	0.990	0.050	33.42	-103.42
55-64 (n = 406)	33.03	10	0.974	0.974	0.075	38.38	133.96
65-74 (n = 375)	40.72	10	0.959	0.959	0.091	38.46	-342.90
≥ 75 (n = 110)	17.89	10	0.955	0.955	0.085	41.12	451.13

Figures

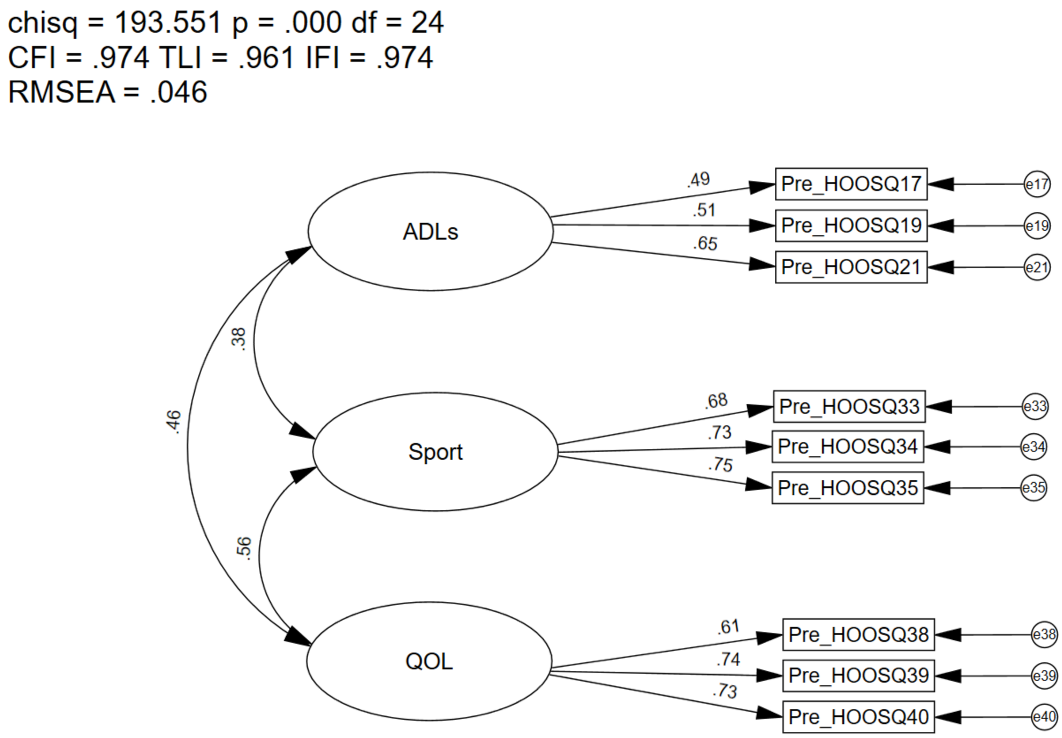
Figure 5

Confirmatory Factor Analysis of the 40-item HOOS

Chisq = Chi Square (χ^2); df = degrees of freedom, p = alpha level; CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; IFI = Bollen's Incremental Fit Index; RMSEA = Root Mean Square Error of Approximation; SRMR = Standardized Root Mean Square Residual.

Figure 6

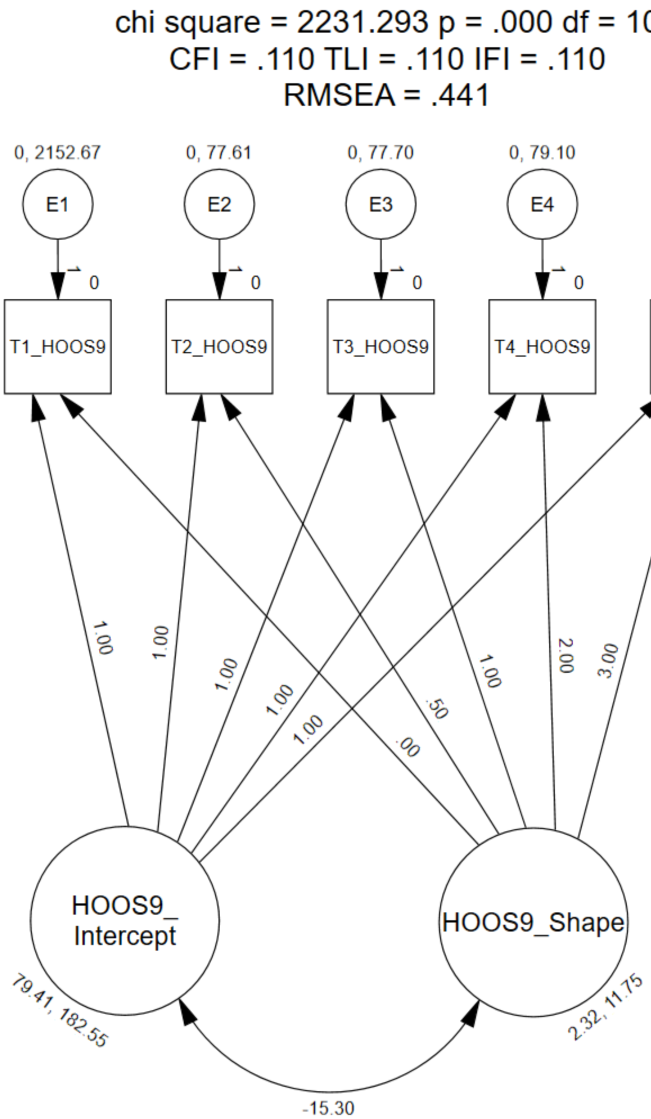
Confirmatory Factor Analysis in Sample n2 of the Modified 9-item HOOS



Chisq = Chi Square (χ^2); df = degrees of freedom, p = alpha level; CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; IFI = Bollen's Incremental Fit Index; RMSEA = Root Mean Square Error of Approximation; SRMR = Standardized Root Mean Square Residual.

Figure 7

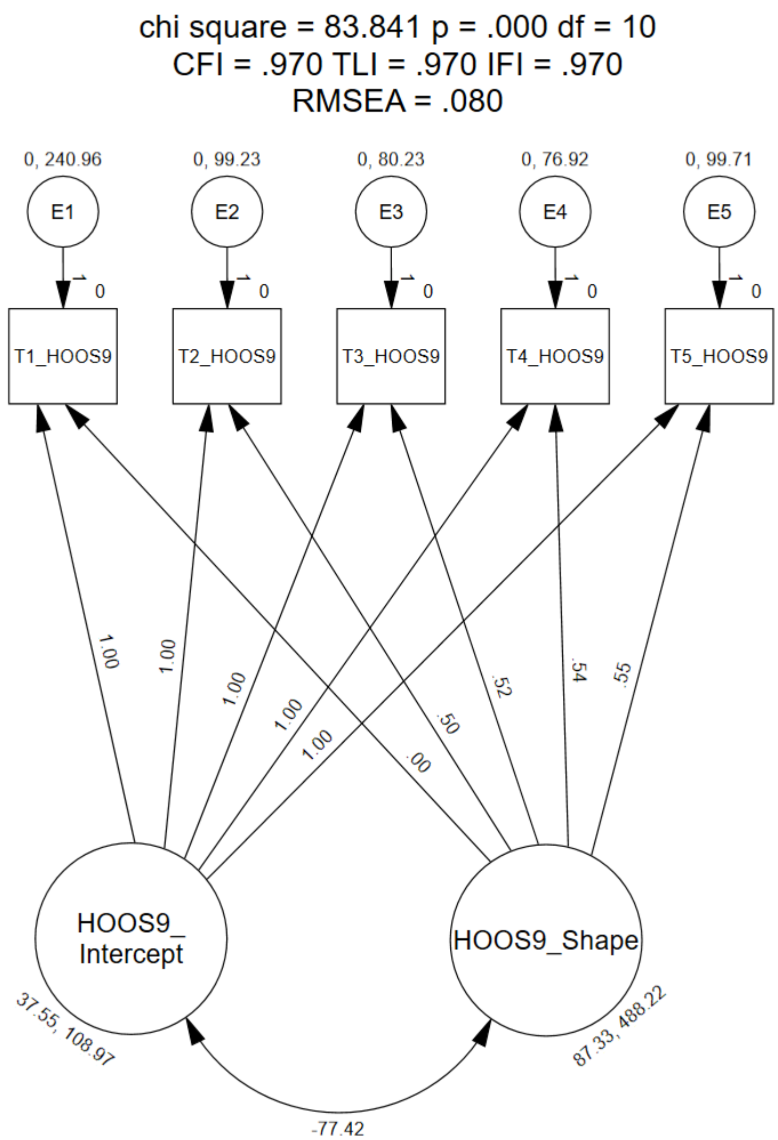
Linear Latent Growth Model of the 9-item HOOS



Chisq = Chi Square (χ^2); df = degrees of freedom, p = alpha level; CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; IFI = Bollen's Incremental Fit Index; RMSEA = Root Mean Square Error of Approximation.

Figure 8

Exploratory Latent Growth Model of the 9-item HOOS



Chisq = Chi Square (χ^2); df = degrees of freedom, p = alpha level; CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; IFI = Bollen's Incremental Fit Index; RMSEA = Root Mean Square Error of Approximation.

Appendix A: Institutional Review Board Letter and Data Use Agreement

CEDARS-SINAI

Office of Research Compliance and Quality Improvement, 6500 Wilshire Blvd., Suite 1800, Los Angeles, CA 90048

IRB APPROVAL NOTICE

June 23, 2020

Dear MICHAEL BANFFY:

On 6/23/2020, the IRB reviewed and approved the following submission:

Type of Submission:	Initial Study
Title of Submission:	STUDY00000841: HOOS Validation Study
Protocol Title:	Evaluating the measurement properties and factorial validity of the Hip Injury and Osteoarthritis Outcome Score (HOOS)
IRB Protocol ID:	STUDY00000841
Investigator:	MICHAEL BANFFY
Funding:	Name: Internal CSMC Funding
IRB Review Level:	Expedited
Approval Effective Date:	6/23/2020
Approval Expiration Date, if applicable:	
Documents Reviewed:	• Data Extraction Sheet , Category: Other; • IRB Protocol, Category: IRB Protocol;

If an expiration date is displayed above, a continuing review must be submitted at least 60 days in advance of this date.

If no expiration date is displayed above, this minimal risk study will not require annual continuing review submissions.

In conducting this research, you are required to follow the IRB approved protocol and all applicable IRB Policies and Procedures.

Agreement ID:

FDP Data Transfer and Use Agreement ("Agreement")

Provider: Cedars-Sinai and its Affiliates/Cedars-Sinai Kerlan-Jobe Institute	Recipient: University of Idaho
Provider Scientist Name: Micheal Banffy Director, Orthopedic Sports Medicine Email: Michael.Banffy@cskerlanjobe.org	Recipient Scientist Name: Russell Baker Email: russellb@uidaho.edu
Agreement Term Start Date: Date of last signature below End Date: Three (3) Years after the Start Date	Project Title: HOOS Validation Study & KOOS Validation Study Attachment 2 Type: De-identified Data about Human Subjects

Terms and Conditions

1) Provider shall provide the data set described in Attachment 1 (the "Data") to Recipient for the research purpose set forth in Attachment 1 (the "Project"). Provider shall retain ownership of any rights it may have in the Data, and Recipient does not obtain any rights in the Data other than as set forth herein.

2) If applicable, reimbursement of any costs associated with the preparation, compilation, and transfer of the Data to the Recipient will be addressed in Attachment 1.

3) Recipient shall not use the Data except as authorized under this Agreement. The Data will be used solely to conduct the Project and solely by Recipient Scientist and Recipient's faculty, employees, fellows, students, and agents ("Recipient Personnel") and Collaborator Personnel (as defined in Attachment 3) that have a need to use, or provide a service in respect of, the Data in connection with the Project and whose obligations of use are consistent with the terms of this Agreement (collectively, "Authorized Persons").

4) Except as authorized under this Agreement or otherwise required by law, Recipient agrees to retain control over the Data and shall not disclose, release, sell, rent, lease, loan, or otherwise grant access to the Data to any third party, except Authorized Persons, without the prior written consent of Provider. Recipient agrees to establish appropriate administrative, technical, and physical safeguards to prevent unauthorized use of or access to the Data and comply with any other special requirements relating to safeguarding of the Data as may be set forth in Attachment 2.

5) Recipient agrees to use the Data in compliance with all applicable laws, rules, and regulations, as well as all professional standards applicable to such research.

6) Recipient is encouraged to make publicly available the results of the Project. Before Recipient submits a paper or abstract for publication or otherwise intends to publicly disclose information about the results of the Project, the Provider will have thirty (30) days from receipt to review proposed manuscripts and ten (10) days from receipt to review proposed abstracts to ensure that the Data is appropriately protected. Provider may request in writing that the proposed publication or other disclosure be delayed for up to thirty (30) additional days as necessary to protect proprietary information.

Appendix B: Hip Disability and Osteoarthritis Outcome Score Joint Replacement

INSTRUCTIONS
This survey asks for your view about your hip. This information will help us keep track of how you feel about your hip and how well you are able to do your usual activities.

Answer every question by ticking the appropriate box, only one box for each question. If you are unsure about how to answer a question, please give the best answer you can.

PAIN

What amount of hip pain have you experienced the **last week** during the following activities?

1. Going up or down stairs
   None ☐ Mild ☐ Moderate ☐ Severe ☐ Extreme ☐

2. Walking on an uneven surface
   None ☐ Mild ☐ Moderate ☐ Severe ☐ Extreme ☐

FUNCTION, DAILY LIVING
The following questions concern your physical function. By this we mean your ability to move around and to look after yourself. For each of the following activities please indicate the degree of difficulty you have experienced in the **last week** due to your hip.

3. Rising from sitting
   None ☐ Mild ☐ Moderate ☐ Severe ☐ Extreme ☐

4. Bending to floor/pick up an object
   None ☐ Mild ☐ Moderate ☐ Severe ☐ Extreme ☐

5. Lying in bed (turning over, maintaining hip position)
   None ☐ Mild ☐ Moderate ☐ Severe ☐ Extreme ☐

6. Sitting
   None ☐ Mild ☐ Moderate ☐ Severe ☐ Extreme ☐

Appendix C: Hip Disability and Osteoarthritis Outcome Score

Name: ____

INSTRUCTIONS: This survey asks for your view about your hip. This information will help us keep track of how you feel about your hip and how well you are able to do your usual activities. Answer every question by ticking the appropriate box, only one box for each question. If you are uncertain about how to answer a question, please give the best answer you can.

SYMPTOMS
These questions should be answered thinking of your hip symptoms and difficulties during the **last week.**

1. Do you feel grinding, hear clicking or any other type of noise from your hip?

Never	Rarely	Sometimes	Often	Always
☐	☐	☐	☐	☐

2. Difficulties spreading legs wide apart

Never	Rarely	Sometimes	Often	Always
☐	☐	☐	☐	☐

3. Difficulties to stride out when walking

Never	Rarely	Sometimes	Often	Always
☐	☐	☐	☐	☐

STIFFNESS
The following questions concern the amount of joint stiffness you have experienced during the **last week** in your hip. Stiffness is a sensation of restriction or slowness in the ease with which you move your hip joint.

4. How severe is your hip joint stiffness after first wakening in the morning?

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

5. How severe is your hip stiffness after sitting, lying or resting **later in the day?**

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

PAIN

6. How often is your hip painful?

Never	Monthly	Weekly	Daily	Always
☐	☐	☐	☐	☐

What amount of hip pain have you experienced the **last week** during the following activities?

7. Straightening your hip fully

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

8. Bending your hip fully

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

9. Walking on a flat surface

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

10. Going up or down stairs

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

11. At night while in bed

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

12. Sitting or lying

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

13. Standing upright

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

14. Walking on a hard surface (asphalt, concrete, etc.)

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

15. Walking on an uneven surface

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

FUNCTION, DAILY LIVING

The following questions concern your physical function. By this we mean your ability to move around and to look after yourself. For each of the following activities please indicate the degree of difficulty you have experienced in the **last week** due to your hip.

16. Descending stairs

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

17. Ascending stairs

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

18. Rising from sitting

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

19. Standing

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

20. Bending to floor/pick up an object

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

21. Walking on flat surface

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

22. Getting in/out of car

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

23. Going shopping

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

24. Putting on socks/stockings

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

25. Rising from bed

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

26. Taking off socks/stockings

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

27. Lying in bed (turning over, maintaining hip position)

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

28. Getting in/out of bath

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

29. Sitting

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

30. Getting on/off toilet

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

31. Heavy domestic duties (moving heavy boxes, scrubbing floors, etc)

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

32. Light domestic duties (cooking, dusting, etc)

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

FUNCTION, SPORTS AND RECREATIONAL ACTIVITIES

The following questions concern your physical function when being active on a higher level. The questions should be answered thinking of what degree of difficulty you have experienced during the **last week** due to your hip.

33. Squatting

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

34. Running

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

35. Twisting/pivoting on your injured hip

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

36. Walking on uneven surface

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

QUALITY OF LIFE

37. How often are you aware of your hip problem?

Never	Monthly	Weekly	Daily	Always
☐	☐	☐	☐	☐

38. Have you modified your lifestyle to avoid potentially damaging activities to your hip?

Not at all	Mildly	Moderately	Severely	Totally
☐	☐	☐	☐	☐

39. How much are you troubled with lack of confidence in your hip?

Not at all	Mildly	Moderately	Severely	Extremely
☐	☐	☐	☐	☐

40. In general, how much difficulty do you have with your hip?

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐