

Normal population reference values for the Oxford and Harris Hip Scores – electronic data collection and its implications for clinical practice

James M. McLean^{1,2}, Jacob Cappelletto¹, Jock Clarnette¹, Catherine L. Hill³, Tiffany Gill³, Daniel Mandziak¹, Jordan Leith²

¹ University of Adelaide Centre for Orthopaedic and Trauma Research, Royal Adelaide Hospital, Adelaide - Australia

² Department of Orthopaedics, University of British Columbia, Vancouver - Canada

³ Discipline of Medicine, Faculty of Health Sciences, University of Adelaide, Adelaide - Australia

ABSTRACT

Background: The aim of this study was to assess whether the Harris Hip Score (HHS) and the Oxford Hip Score (OHS) were comparable in normal, healthy, pathology-free individuals of different age, gender, ethnicity, handedness and nationality. The purpose of this study was to establish normal population values for the HHS and OHS using an electronic data collection system.

Methods: 317 Australian and 310 Canadian citizens with no active hip pain, injury or pathology in the ipsilateral hip corresponding to their dominant arm, were evaluated. Participants completed an electronically-administered questionnaire and were assessed clinically. Chi-square tests, Fisher's exact test and Poisson regression models were used where appropriate, to investigate the association between hip scores, ethnicity, nationality, gender, handedness and age.

Results: There was a statistically significant association between the OHS and age ($p < 0.0001$) and the HHS and age ($p = 0.0006$); demonstrating that as age increased, normal hip scores decreased. There was no statistically significant association between the HHS and gender ($p = 0.1389$); or HSS and nationality, adjusting for age ($p = 0.5698$) and adjusting for gender ($p = 0.6997$). There was no statistically significant association between the OHS and gender ($p = 0.1350$). Australians reported a statistically significant 4.2% higher overall OHS value compared to Canadians ($p = 0.0490$). There was no statistically significant association between the OHS and nationality in age groups 18-79 years. Participants >80 years reported a statistically significant association between the OHS and nationality ($p < 0.0001$).

Conclusions: Studies using an electronic control group should consider differences in gender, age, ethnicity and nationality when using the HHS and OHS to assess patient outcomes. This study has established an electronic, normal control group for studies using the HHS and OHS. When using the OHS, the control group should be sourced from the same country of origin. When using the HHS, the control group should be sourced from a pre-established control group within a database, without necessarily being sourced from the same country of origin.

Keywords: Arthroplasty, Harris, Hip, Outcome, Oxford, Reference values

Introduction

Advances in total hip arthroplasty (THA) have seen an increase in the survivability of implants and a decrease in cumulative revision rates over the past 30 years. The long-term results

of THA require longitudinal patient and implant assessments. Accurate interpretation of these long term studies requires an understanding of not only surgical and implant factors, but also patient factors that may change over time.

Pre- and post-operative patient reported outcomes measures (PROMs) can be used to measure the severity of a patient's symptoms and level of function. They can be important tools in assessing a patient's suitability for surgery, expected outcome and post-operative recovery. PROM data can now be collected using computer-based, electronic data collection systems, that allow for quicker data collection, automated data processing, and minimal clinician input (1).

Several THA PROM clinical scores have been described, validated and compared (2-9). Each assessment tool has its relative strengths and benefits; as well as its weaknesses. No single assessment tool has reported consistent superiority

Accepted: August 23, 2016

Published online: November 15, 2016

Corresponding author:

James M. McLean
Discipline of Orthopaedics and Trauma
University of Adelaide
North Terrace
5005 Adelaide, SA, Australia
james.mc_lean@adelaide.edu.au

over another and the choice of which one to use generally is determined by patient population, pathology, investigator preference and resource management (3, 5, 7-9). The Harris Hip Score (HHS) combines subjective PROM patient inputs with objective, clinician-derived inputs to derive a score (10). The HHS is widely used (5, 6, 10) and has been shown to have acceptable reliability and construct validity (11). The Oxford Hip Score (OHS) includes PROMs of pain and function (12). It has been shown to have excellent reliability and construct validity (13, 14).

Accurate interpretation of long-term THA studies requires an understanding of patient, surgical and implant factors that may change over time (15). A perfect hip score may not reflect a realistic goal, as an accurate interpretation of a patient's score requires a comparison with an age- and gender-matched group of individuals who have not had a THA (7, 15).

The purpose of this study was to establish normal population values for the HHS and OHS using an electronic data collection system.

Our hypothesis was that there is no difference in the HHS and OHS values in a normal population when comparing age, gender, ethnicity, handedness and different nationalities.

Methods

Independent Ethics Board approval was granted from each Institution involved. From November 2014 to May 2015, healthy volunteers were recruited from a variety of sources, including Drivers Licensing Offices; Medical Outpatients Facilities; and various community centres (sporting, childcare, recreation, library and senior's activity facilities). There were no study advertisements or incentives and participants were not paid for their involvement.

Adult participants were approached if they were fluent in English, and were Australian or Canadian citizens. The inclusion criteria included no active hip pathology in the hip corresponding to their dominant arm. Potential participants self-reported a history of hip pain or hip pathology; no medical charts or radiographs were reviewed to categorize asymptomatic participants.

Exclusion criteria included: cognitive impairment; a history of inflammatory or hip arthritis; significant lumbar spine problems that interfered with their function; active hip pathology; hip arthroplasty; or hip surgery within the past 3 years. A history of inactive hip pathology, including previous surgery, was recorded. A history of active knee/ankle/foot pathology was recorded.

Participants self-administered 20 questions (OHS 12 questions; HHS 8 questions) using a web-based data collection tool (OBERD, Universal Research Solutions), on an electronic mobile device (electronic tablet or laptop computer). This method enabled minimal data handling by the recruiters, ensuring that the investigators were partially-blinded to the participants' results. An option to provide feedback was given.

Participants' range of motion (RoM) was recorded. A single, highly experienced observer performed all assessments in Canada. In Australia, 2 observers with less experience performed all assessments. Interobserver variability correlation was performed.

Primary outcome measures

Harris Hip Score (HHS)

The HHS is a 13-item patient/clinician report of pain (44-points); function (47-points); deformity (4-points); and ROM (5-points) (10). A visual analogue scale is used and then scaled to a 100-point sum (maximum perfect score = 100).

Oxford Hip Score (OHS)

The OHS is a 12-item patient report of pain (6-items) and function (6-items) (12). Each item is scored from 0-4 points (maximum perfect score = 48) (16).

Statistical analysis

A power calculation was performed to determine the sample size necessary to detect a clinically significant difference in hip scores of 20% at a power of 80% and an alpha value of 0.05 ($n = 596$). Analyses were performed using the IBM SPSS V.20 statistical package and SAS.9.3 (SAS Institute Inc.).

Associations between nationality and age, gender, handedness and ethnicity were investigated using chi-square and Fisher's exact test where appropriate. Poisson regression models were used to investigate the association between hip scores and these variables. Linear regression was not performed because residuals from a linear model were very left-skewed, as were the residuals using a logarithmic transform of the outcome variable. Hip scores were therefore considered to be counts. Poisson regressions were performed and ranged from 0.0124 to 3.1994. CI was set at 95% for 2-way mixed effects model and absolute agreement. Initially nationality cohort and all confounders were included in a multivariable Poisson regression model for each hip score outcome variable. Backwards stepwise elimination was then performed until all covariates had a p value < 0.2 .

Results

The demographics of the cohorts are presented in Tables I and II.

Overall 2.6% of Canadian and 3.8% of Australian participants felt that 20 questions were too many. These respondents would have preferred to answer 8 (range 0-10) or 10 questions (range 0-17), respectively.

The incidence of participants reporting a *history of an inactive hip problem* is presented in Table II. 1 Canadian and no Australians reported having had non-arthroplasty hip surgery that was performed more than 3 years ago. Participants with a *history of an inactive hip problem* had a mean OHS value 9% lower ($p = 0.0865$; IRR = 1.09; 95% CI, 0.99-1.17); and a mean HHS value 6% lower ($p = 0.0478$; IRR = 1.06; 95% CI 1.0006-1.12), than those who reported no such history. A statistically significant association was borderline for HHS ($p = 0.0478$).

The incidence of participants reporting an *active knee/ankle/foot problem* is presented in Table II. Participants with an *active knee/ankle/foot problem* reported a statistically significant 9% lower OHS value ($p < 0.0001$; IRR = 1.09, 95% CI,

TABLE I - A comparison of ethnicity of the 2 international cohorts. A statistical difference was demonstrated when comparing ethnicity ($p < 0.0001$). Due to the relatively low numbers recorded in some ethnic groups, no statistically significant comparisons could be made between the individual ethnic groups

Ethnicity	Australia (n = 317)	Canada (n = 310)	Total (n = 627)
Asian Indian	3 (0.9%)	18 (5.8%)	21 (3.3%)
Black or African American	1 (0.3%)	3 (1%)	4 (<1%)
Caucasian	302 (95%)	221 (71%)	523 (83.4%)
Chinese	2 (0.6%)	32 (10.3%)	34 (5.4%)
Filipino	1 (0.3%)	7 (2.2%)	8 (1.2%)
Indigenous	0	1 (0.3%)	1 (<1%)
Middle Eastern	2 (0.6%)	14 (4.5%)	16 (2.6%)
Other Asian	6 (2%)	14 (4.5%)	20 (3.2%)

1.05-1.14); and 9% lower HHS value ($p < 0.0001$; IRR = 1.09, 95% CI, 1.06-1.12).

Harris Hip Score - clinician objective component

82 participants (12.8%) did not score the potential 9/9 for the clinician-assessed objective component. Of these participants, the average score was 8.25/9 (Range: 4.75-8.85); 79/82 had a total RoM 70°-100° (representing a loss of <1 point; range 0.25-0.85); 3/82 had a leg length discrepancy >1.5 inches (representing a loss of 4 points). No participants had a fixed flexion deformity >30°; <20° abduction; or <15° of internal or external rotation.

Harris Hip Score - total

There was a statistically significant association between the HHS and age ($p = 0.0006$; IRR = 0.9991, 95% CI: 0.9986, 0.9996). For every 1-year increase in age, the mean HHS value decreased by 0.1% (Fig. 1, Tab. III).

There was no statistically significant association between HHS and gender ($p = 0.1389$); handedness ($p = 0.5564$); or nationality (adjusting for age ($p = 0.5698$); and adjusting for gender ($p = 0.6997$)). Australians reported HHS values 2.7%

TABLE II - A comparison of demographics of the 2 international cohorts

	Australian cohort	Canadian cohort	Total	Comparing Australia and Canadian cohorts
Male	159 (50.2%)	154 (50.0%)	315 (50.1%)	$p = 0.9684$
Female	158 (49.8%)	156 (50.0%)	314 (49.9%)	
Left	33 (10%)	25 (8%)	57 (17.4%)	$p = 0.1810$
Right	284 (90%)	285 (92%)	570 (82.6%)	
Age <30	36 (11.4%)	29 (9.3%)	65 (10.3%)	$p = 0.9772$
Age 30-39	34 (10.7%)	34 (10.9%)	68 (10.8%)	
Age 40-49	51 (16.1%)	53 (17.0%)	104 (16.5%)	
Age 50-59	72 (22.7%)	74 (23.7%)	146 (23.2%)	
Age 60-69	71 (22.4%)	70 (22.4%)	141 (22.4%)	
Age 70-79	33 (10.4%)	37 (11.9%)	70 (11.1%)	
Age 80+	20 (6.3%)	15 (4.8%)	35 (5.6%)	
Privately insured	160 (50.5%)	0		
Publically insured	157 (49.5%)	310		
Average age	53 years (range 18-90)	53 years (range 18-94)	53 years (range 18-94)	$p = 0.9772$
Patient reported a history of an inactive (previous) hip problem	4 (1.3%)	10 (3.2%)	14 (2.2%)	$p = 0.1108$ (OR = 0.39, 95% CI: 0.12, 1.24)
Patient reported a history of an active knee/ankle/foot problem	9 (2.8%)	45 (14.4%)	54 (8.6%)	$p < 0.0001$ (OR = 0.17, 95% CI: 0.08, 1.36)

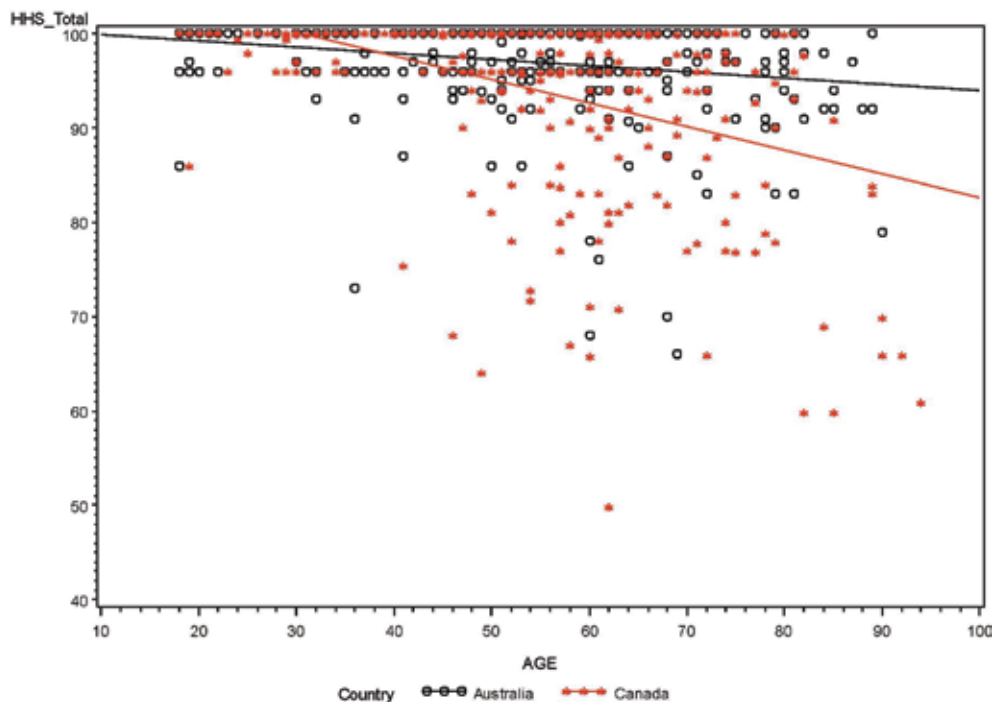


Fig. 1 - A scattergram of the Harris Hip Score versus age. Maximum score = 100 points. Red - Canada; Black - Australia.

TABLE III - Harris Hip Score (HHS - maximum 100 points)

Participant age group	Australia Ave	Canada Ave	Combined
<30	98.6958	98.9871	98.6501
30-39	97.6471	99.2309	98.4273
40-49	98.3113	96.8594	97.5887
50-59	97.7469	94.5135	96.1169
60-69	95.9384	92.2268	94.0772
70-79	95.5455	90.7230	93.0674
80+	94.1500	79.6983	87.7525
Overall	96.5055	93.5727	95.64

Average HHS scores for Australian and Canadian cohorts, and the combined Australian and Canadian cohorts. There was no statistically significant association between HHS and nationality, adjusting for age ($p = 0.5698$); and adjusting for gender ($p = 0.4888$). Ave = Average; Combined = Australian and Canadian participants combined in that age group.

greater than Canadians, which was not statistically significant (Tab. III; $p = 0.5698$; IRR = 1.0055, 95% CI, 1.018-1.067).

Oxford Hip Score

There was a statistically significant association between the OHS and HHS ($p < 0.0001$; IRR = 1.017, 95% CI: 1.015, 1.020). For every one unit increase in the OHS, the HHS value increased by 1.7%.

There was a statistically significant association between OHS and age, adjusting for nationality ($p < 0.0001$; IRR = 0.9976, 95% CI: 0.9969, 0.9983). For every 1-year increase in age, the mean OHS value decreased by 0.24% (Tab. IV).

There was no statistically significant association between OHS and gender ($p = 0.1350$) or handedness ($p = 0.4301$).

There was a statistically significant association between the OHS and nationality (Fig. 2); adjusting for age ($p = 0.0490$) and adjusting for gender ($p = 0.0003$). Australians reported a mean OHS value 4.2% greater than Canadians (Tab. IV; $p = 0.0490$; IRR = 1.042, 95% CI, 1.018-1.067).

Discussion

It is an important goal to differentiate normal, age-related changes in function, from those changes associated with THA wear, fatigue or failure (17, 18). An important step towards



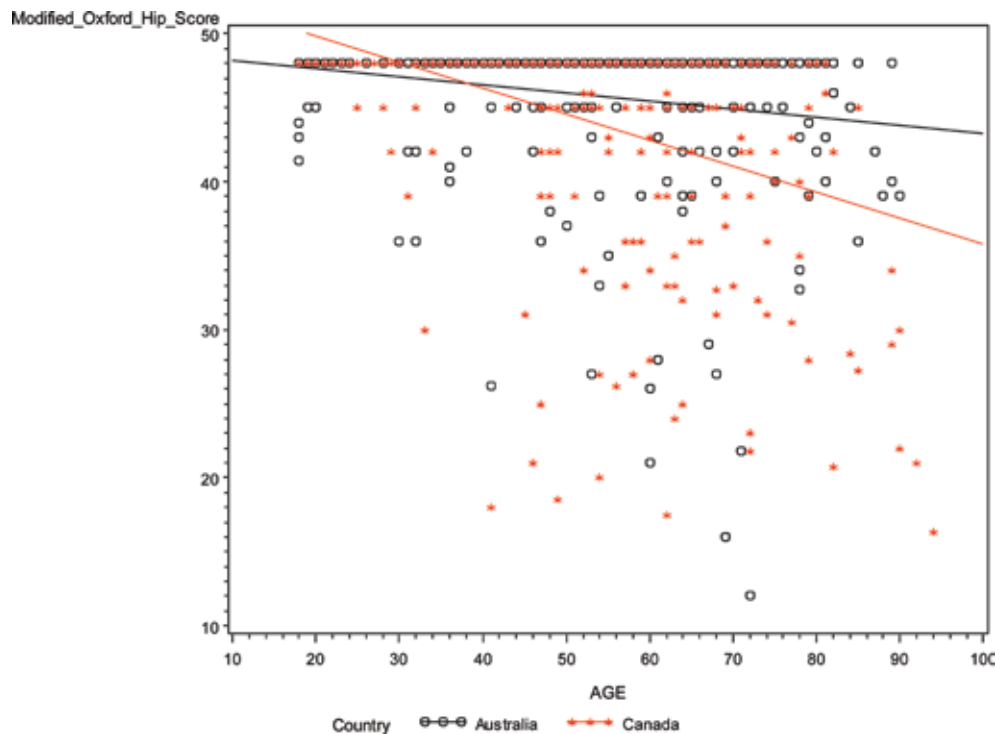


Fig. 2 - A scattergram of the Oxford Hip Score versus age. Maximum score = 48 points. Red - Canada; Black - Australia.

TABLE IV - Oxford Hip Score (OHS - maximum 48 points)

Participant age group	Australia Ave	Canada Ave	Combined	P value
<30	47.4015	47.5862	47.3613	0.9145
30-39	46.2353	46.9412	46.5771	0.6698
40-49	46.6702	44.6518	45.6499	0.1278
50-59	46.6389	45.0295	45.8259	0.1510
60-69	44.8169	42.7455	43.7713	0.0631
70-79	43.8044	41.1720	42.4558	0.0914
80+	44.4500	33.7152	39.7160	<0.0001
Overall	45.3882	43.4493	44.88	0.0490

Average OHS scores for Australian and Canadian cohorts, and the combined Australian and Canadian cohorts. There was a statistically significant association between the OHS and nationality, adjusting for age ($p = 0.0490$) and adjusting for gender ($p = 0.0003$). However, there was no statistically significant difference between the international cohorts when comparing specific age groups in participants <79 years of age. Participants >80 years of age had a larger variation in score and a statistically significant difference was observed between the international cohorts. Ave = Average; Combined = Australian and Canadian participants combined in that age group.

this goal is establishing a reference database for individuals without hip disease, so that we can effectively evaluate the efficacy of THA patients on a longitudinal basis.

In this study, data were collected electronically from 2 normal, distinct, remote, Westernised populations of different countries, that were representative of their local populations (19, 20). To our knowledge, this has not been investigated previously. The higher proportion of persons of European descent (Caucasians by default) represented in the Australian cohort is consistent with that reported by the Australian Bureau of Statistics (19). The higher proportions of

Chinese, Middle Eastern and Asian Indians represented in the Canadian cohort, is consistent with that reported by Statistics Canada (20). Although there was a difference observed between the cohorts in regard to ethnicity, the numbers were too small to allow for any statistical assessment.

We chose to assess the OHS because it contains subjective-only reports hip function (12) and has the potential advantage over the HHS of being administered remotely – without the need for a face-to-face interaction with the participant. We chose to assess the HHS because it contains both subjective and objective components. We wanted to assess



whether the potential benefits of using an electronically-administered assessment tool were negated by the need for a clinical assessment by a skilled observer, which requires allocated time, appropriate outpatient facilities and a face-to-face interaction. We also wanted to directly compare these assessment tools to determine if a subjective-only tool had any advantage over a combined subjective/objective tool.

This study demonstrated that OHS values differed between the international cohorts. However, when the age groups were assessed individually, no difference was found between cohorts for participants <79 years (Tab. IV). The greatest variation in OHS values was recorded in the ≥80 age group, with some respondents recording OHS values of 48/48 (even up to 90 years (OHS range 13-48/48). It is not surprising, given this large variation in respondent's subjective reports of hip function, that a statistically significant difference was identified between the two cohorts in the ≥80 age group. Overall, this study suggests that normal OHS values are comparable between countries for individuals <80 years, and highly variable in individuals ≥80 years, making comparisons in this older age group less reliable.

Care should be taken when interpreting these data and applying generalisations to different populations. Specifically, THA patients ≥80 years, should be compared to a gender-matched control group sourced from their same country of origin.

This study demonstrated that HHS values were comparable between the national cohorts. This suggests that future HSS studies can be performed using a combined control group, without necessarily needing to be sourced from the same country of origin as the proposed study. Further studies need to be completed to determine whether this principle applies to other countries that use this same electronic database, particularly the United States and Great Britain.

An inverse relationship was observed between age and clinical score. This finding is not surprising, given the age-related changes that occur over time, as well as the accumulated medical and surgical comorbidities that can affect lower limb function. The large variation in OHS reported in individuals >80 years, suggesting that there are likely many other determinants of health and function that may influence the subjective score reported, and possibly the accuracy of "normal" values in this age group.

This study did not report an association between a *history of inactive hip pathology* and OHS; but did report an association with HHS. This finding is difficult to interpret and may reflect a selection or reporting bias. It would have seemed logical that a participant with a *history of inactive hip pathology* would have a lower overall hip score.

This study reported an association between a *history of active knee/ankle/foot pathology* and clinical score. There is significant overlap in the functional questions contained in the assessment tools, suggesting that these functional tools may not represent hip-specific PROMs and may fail to discriminate a primarily hip-source of pathology from other sources of lower limb incapacity. This should be considered when using these tools in patients with multiple, concurrent lower limb pathologies.

To our knowledge, our study represents the largest database of normal HHS and OHS values reported in the literature.

Other researchers have recorded normal values for other musculoskeletal assessment tools, but few have approached the numbers collected in this study, with most collecting less than 150 participants (21), or limited to young, active individuals (22).

Lieberman et al (15) reported on 184 individuals >55 years and established normal HHS values for this group. However, their questionnaires were administered by telephone and no clinical assessments were performed. In their methods, they assigned all participants 9/9 points for objective measurements when calculating the HHS (15). In our study, 12.8% participants did not score the complete 9/9 assigned by Lieberman et al (15). Of these, the average number of points allocated was 8.25/9 points (range 4.75-8.85), with only 3/627 scoring less than 6/9 (all 3 losing 4 points secondary to a leg length discrepancy).

We committed resources to collecting objective clinical data to complete the HHS. This study demonstrated that the collection of objective data contributed to <1/100-point difference in 12.3% and up to 4.125/100 points in <0.5% of participants. This important finding led us to re-evaluate the importance of collecting objective data for calculating the HSS. As resource management and cost justification is becoming more of a focus for our Institutions, consideration should be given to a hip PROM tool that is equivalent to the HHS but does not require an objective assessment component. Other investigators have also questioned the clinical applicability of the HHS and have recommended other hip PROMs in its place (11, 23, 24).

A subjective-only hip PROM assessment tool has several advantages. One of these advantages includes the ability to administer questionnaires remotely, negating the need for patients to be reviewed by a clinician, thereby increasing their cost-effectiveness. They can also be automatically administered, quickly and easily, with reproducible results. However, electronically-administered questionnaires have a lower response rate and require respondents to be computer savvy; an assumption that may not be correct for all members of the public, especially the elderly THA patient population.

Byrd et al (25) introduced a modified HHS (mHHS) to assess the outcomes of young patients following arthroscopic hip debridement. In their description, the 9-point objective component was omitted and the subjective components (maximum 91 points) were multiplied by 1.1 to give a total maximum score of 100 (25). To our knowledge, the mHHS has not been tested for content validity or reliability (3, 8, 26). Further research needs to be done to compare electronically administered subjective-only and combined subjective/objective hip PROMs with a pathological group. This study has established the control group for such a study.

The current study has important limitations that should be considered when interpreting the results. As with any observational study, there is the potential for selection bias, particularly when there is no randomisation. The primary benefits of randomisation are the elimination of both conscious and unconscious bias associated with the selection of a participant. Although individuals were approached randomly in this study, no specific randomisation method of participant identification was employed. Another potential source of selection bias involves the use of electronic questionnaires,

where participants may have declined to be involved due to the technology. Anecdotally, several elderly participants were initially reluctant to be involved, but agreed to participate with an assessor helping complete the electronic questionnaires. This may have introduced interviewer bias.

Participants with a history of a prior hip injury may have chosen not to participate in the study, citing that their hip was not “normal”. Although we chose to exclude participants with *active hip disease*, we did include participants with a *history of a previous hip problem* that “no longer bothered them”. As this is a purely subjective report, it is possible that some of those individuals who had a prior hip problem may have only minor functional incapacities, and should have been included in the study. There was no difference reported between the international cohorts in relation to a *history of inactive hip pathology*. As no x-rays were taken to confirm whether participants had asymptomatic degenerative hip disease, it is possible that some of these individuals were included in the cohorts.

There was a statistically significant difference reported between the cohorts when comparing a *history of an active knee/ankle/foot problem*. It is possible that a selection bias may have contributed to this finding. As this study reported an association between a *history of an active knee/ankle/foot pathology* and clinical score, the higher proportion of Canadian participants who reported problems, may explain the overall lower Canadian HHS and OHS values reported in this study.

Conclusions

Differences in age, gender, ethnicity and nationality should be taken into consideration when using the HHS and OHS to assess patient outcomes. A larger sample size would need to be collected to assess for subtle differences in ethnicity.

Studies using the OHS and an electronic, pre-established control group, should be sourced from the same country of origin and be age- and gender-matched. Future electronic database-derived studies that use the HSS, can utilize the combined, pooled control group as a comparative group, without necessarily needing to be sourced from the same country of origin as the proposed study. Further studies need to be completed to determine whether this principle applies to other countries that use this same electronic database.

Acknowledgement

The authors would like to thank:

- (i) Mami Okada, Research Assistant, University of British Columbia, for her help in the establishment of the study in Canada.
- (ii) Tara-Louise McLean, Research Assistant to Dr James McLean, for her help in the establishment of the study in Australia and Canada.
- (iii) Suzanne Edwards, Statistician, Data Management and Analysis Centre at the University of Adelaide, for her help with the statistical analysis of the collected data.

Disclosures

Financial support: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. No sponsorship was sought. The investigators independently paid

for a license to use the research software and database. The individual research tool registrations were paid for by OBERD Pty Ltd and are included as part of the licensing agreement. Conflict of Interest: None.

References

1. Griffiths-Jones W, Norton MR, Fern ED, Williams DH. The Equivalence of Remote Electronic and Paper Patient Reported Outcome (PRO) Collection. *J Arthroplasty*. 2014;29(11):2136-2139.
2. Thorborg K, Tjissen M, Habets B, et al. Patient-Reported Outcome (PRO) questionnaires for young to middle-aged adults with hip and groin disability: a systematic review of the clinimetric evidence. *Br J Sports Med*. 2015;49(12):812.
3. Hinman RS, Dobson F, Takla A, O'Donnell J, Bennell KL. Which is the most useful patient-reported outcome in femoroacetabular impingement? Test-retest reliability of six questionnaires. *Br J Sports Med*. 2014;48(6):458-463.
4. 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.
5. Collins NJ, Roos EM. Patient-reported outcomes for total hip and knee arthroplasty: commonly used instruments and attributes of a “good” measure. *Clin Geriatr Med*. 2012;28(3):367-394.
6. Ware J Jr, Kosinski M, Keller SDA. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Med Care*. 1996;34(3):220-233.
7. Alviar MJ, Olver J, Brand C, Hale T, Khan F. Do patient-reported outcome measures used in assessing outcomes in rehabilitation after hip and knee arthroplasty capture issues relevant to patients? Results of a systematic review and ICF linking process. *J Rehabil Med*. 2011;43(5):374-381.
8. Kemp JL, Collins NJ, Roos EM, Crossley KM. Psychometric properties of patient-reported outcome measures for hip arthroscopic surgery. *Am J Sports Med*. 2013;41(9):2065-2073.
9. Söderman P, Malchau H, Herberts P. Outcome of total hip replacement: a comparison of different measurement methods. *Clin Orthop Relat Res*. 2001;(390):163-172.
10. Harris WH. Traumatic arthritis of the hip after dislocation and acetabular fractures: treatment by mold arthroplasty. An end-result study using a new method of result evaluation. *J Bone Joint Surg Am*. 1969;51(4):737-755.
11. Söderman P, Malchau H. Is the Harris hip score system useful to study the outcome of total hip replacement? *Clin Orthop Relat Res*. 2001;(384):189-197.
12. Dawson J, Fitzpatrick R, Carr A, Murray D. Questionnaire on the perceptions of patients about total hip replacement. *J Bone Joint Surg Br*. 1996;78(2):185-190.
13. Dunbar MJ, Robertsson O, Ryd L, Lidgren L. Appropriate questionnaires for knee arthroplasty. Results of a survey of 3600 patients from The Swedish Knee Arthroplasty Registry. *J Bone Joint Surg Br*. 2001;83(3):339-344.
14. 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.
15. Lieberman JR, Hawker G, Wright JG. Hip function in patients >55 years old: population reference values. *J Arthroplasty*. 2001;16(7):901-904.
16. Murray DW, Fitzpatrick R, Rogers K, et al. The use of the Oxford hip and knee scores. *J Bone Joint Surg Br*. 2007;89(8):1010-1014.
17. Brinker MR, Lund PJ, Cox DD, Barrack RL. Demographic biases found in scoring instruments of total hip arthroplasty. *J Arthroplasty*. 1996;11(7):820-830.

18. Ilstrup DM, Nolan DR, Beckenbaugh RD, Coventry MB. Factors influencing the results in 2,012 total hip arthroplasties. *Clin Orthop Relat Res.* 1973;(95):250-262.
19. Cultural Diversity in Australia. *Reflecting a Nation: Stories from the 2011 Census.* Canberra: Australian Bureau of Statistics; 2011.
20. Ethnic origins for Canada, provinces and territories *Census of Population.* Calgary: Government of Canada; 2011.
21. Engelberg R, Martin DP, Agel J, Swiontkowski MF. Musculoskeletal function assessment: reference values for patient and non-patient samples. *J Orthop Res.* 1999;17(1):101-109.
22. Cameron KL, Thompson BS, Peck KY, Owens BD, Marshall SW, Svoboda SJ. Normative values for the KOOS and WOMAC in a young athletic population: history of knee ligament injury is associated with lower scores. *Am J Sports Med.* 2013;41(3):582-589.
23. Garellick G, Malchau H, Herberts P. Specific or general health outcome measures in the evaluation of total hip replacement. A comparison between the Harris hip score and the Nottingham Health Profile. *J Bone Joint Surg Br.* 1998;80(4):600-606.
24. Lieberman JR, Dorey F, Shekelle P, et al. Outcome after total hip arthroplasty. Comparison of a traditional disease-specific and a quality-of-life measurement of outcome. *J Arthroplasty.* 1997;12(6):639-645.
25. Byrd JW, Jones KS. Prospective analysis of hip arthroscopy with 2-year follow-up. *Arthroscopy.* 2000;16(6):578-587.
26. Potter BK, Freedman BA, Andersen RC, Bojeskul JA, Kuklo TR, Murphy KP. Correlation of Short Form-36 and disability status with outcomes of arthroscopic acetabular labral debridement. *Am J Sports Med.* 2005;33(6):864-870.