The establishment of an automated, electronic data-collection system for the purpose of qualifying, evaluating and improving post-operative surgical outcomes and evidence-based surgery

James Marcus McLean MS MBBS FRACS FA(Ortho)A

Senior Clinical Lecturer

Discipline of Orthopaedics & Trauma

School of Medicine

University of Adelaide, Australia

April 2019

Thesis by publication submitted for the degree of Doctor of Philosophy, University of Adelaide.

TABLE OF CONTENTS

ABSTRACT

It is an important goal to differentiate normal, age-related changes in joint function from changes associated with symptomatic pathology, surgical intervention or implant wear, fatigue or failure. A significant step toward this goal is the establishment of a joint-specific Patient Recorded Outcome Measures (PROM) reference database for individuals without joint disease, so that clinicians can effectively evaluate the efficacy of treatments in orthopaedic patients on a longitudinal basis.

This thesis has investigated the influence of a range of factors on reported PROM values. Factors assessed were age, gender, ethnicity, handedness (where applicable), nationality, history of previous surgery and coincident adjacent active joint pathology. No other research exists comparing multiple PROMs over multiple body regions within an electronic database across two continents.

This thesis of 2360 participants represents the largest database reported in the literature of orthopaedic PROM values from asymptomatic "normal" individuals. No other study has collated "normal" PROM values from multiple body regions. Nor have previous studies collected PROM scores remotely and electronically via the same research database, across two continents. Few studies have approached the numbers collected in this study.

The PROMs investigated cover the four major joints, for which the majority of literature exists (hip, knee, shoulder & hand/wrist). The collected data has established a database of PROM population reference values for individuals who identify themselves as asymptomatic for the body parts under investigation. It is intended that these pooled values can be used as asymptomatic control cohorts for future studies investigating pathological cohorts.

STATEMENT OF ORIGINALITY

This thesis contains material that has been published in peer-reviewed medical journals.
I certify that this work contains no material, which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I acknowledge that copyright of published works contained within this thesis resides with the copyright holder(s) of those works.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

I have been the first author or principal investigator of the publications that form the main body of the thesis. Co-authors of any of the papers have provided support for papers to be included in this thesis. The individual statements of the contributions of jointly authored papers are in the section titled "Statement of authorship and contribution".

1. **Normal population reference values for the Oxford and Harris Hip Scores – electronic data collection and its implications for clinical practice**
   James M. McLean, Jacob Cappelletto, Jock Clarnette, Catherine L. Hill, Tiffany Gill, Daniel Mandziak, Jordan Leith.
   Hip International. 2017; 27 (4): 389-396

2. **Asymptomatic Population Reference Values for Three Knee Patient-Reported Outcomes Measures – Evaluation of an Electronic Data Collection System and Implications For Future International, Multi-Centre Cohort Studies**
   James M. McLean, Oscar Brumby-Rendell, Ryan Lisle, Jacob Brazier, Kieran Dunn, Tiffany Gill, Catherine L. Hill, Daniel Mandziak, Jordan Leith.
   Archives of Orthopaedics & Trauma Surgery. 2018; 138(5): 611-621.

3. **<u>An International, Multi-Centre Cohort Study Comparing Six Shoulder Clinical Scores in an Asymptomatic Population</u>**

James M. McLean, Daniel Awwad, Ryan Lisle, James Besanko, Donald Shivakkumar, Jordan Leith.

Journal of Shoulder & Elbow Surgery. 2018; 27(2): 306-314.

4. **<u>Asymptomatic Reference Values for the DASH and PRWHE - Electronic Data Collection and Its Clinical Implications</u>**

James M. McLean, Afsana Hasan, Jake Willet, Matthew Jennings, Kimberley Brown, Laura Goodwins, Tom Goetz.

urgery: European Volume. 2018; 43(9): 988-993.

James M. McLean 1st April 2019

DEDICATION

This thesis is dedicated to my partner Tara-Louise for her unreserved support and understanding in the pursuit of my professional goals. Her love and devotion have inspired me to become a more rounded human being. I look forward to sharing a life with her into the future.

To my father Steven, for inspiring my pursuit of knowledge and instilling in me a sense of compassion and responsibility. To my mother Susan for nurturing empathy and kindness. To my parents-in-law, Maxine and Terry, for their unwavering support, generosity and benevolence, irrespective of circumstance. Your example and support has enabled me to follow my dreams, establish a loving family of my own and mature into a proud nurturer.

The challenge shall be for Tara-Louise and me to emulate the support and encouragement they have shown; and to pass these special qualities on to our sons William and Emmett; and other welcomed additions to our family.

ACKNOWLEDGEMENTS

I am indebted to Mr Gregory Bain, whose moral standard and professionalism has inspired me to pursue higher standards both personally and professionally. I thank him for his inspiration and generous support of my research through his teaching and guidance.

I thank Suzanne Edwards, Statistician, Data Management and Analysis Centre at the University of Adelaide, for her help with the statistical analysis of the collected data.
I thank Ms Mami Okada who guided me through the process of preparing and submitting each of the research ethics applications in Canada.

I thank Dr David Findlay for his support of the culmination of my work as a thesis within the university's framework.

I thank Dr Thomas Goetz for his support in setting up the research in Canada to achieve the research goals.

I thank Dr Tiffany Gill and Catherine Hill for their help with the conceptual design and supervision of the project.

PRIZES & AWARDS

Keith Doddridge Prize - “Best paper”

South Australian Hand Surgery Society Annual Scientific Meeting

25 Aug 2017

CHAPTER 1 - BACKGROUND & LITERATURE REVIEW

In an effort to improve patient outcomes, the research conducted in preparation for this thesis has been performed with the consideration below in mind.

For clinical publications, most high-impact journals require outcome data with a minimum of 2 years follow-up; and a control group for comparison.

This thesis is primarily aimed at addressing the second part of this requirement, that being the establishment of a control group for comparison.

The efficacy of orthopaedic surgical outcomes can sometimes be difficult to quantify and evaluate. There is often a spectrum of clinical and surgical results, with individual patients having different thresholds for their interpretation of successful or unsuccessful surgery. Objective outcome measurements such as joint range of motion, hospitalisation cost [(1)], hospital stay [(2)], metres walked, and/or complication(s), which can be easily recorded and tabulated, have been used to qualify the degree of surgical improvement or failure.

Other subjective patient reports of pain and function have also been used to describe a patient's perception of recovery and/or improvement [(3)]. However, these reports often lack homogenous values, thus making comparisons between individuals or different study cohorts difficult.

In an effort to quantify subjective and objective pre-operative and post-operative patient assessments, patient-reported outcomes (PROs), also referred to as Patient-reported outcomes measures (PROMs), were introduced in the 1970s[(4)]. These scoring systems offered the opportunity to increase the comparability of orthopaedic surgical outcome assessments [(5-8)]. This might enable surgeons to more confidently measure the severity of a patient's symptoms and level of function, both in real time and longitudinally. Standardised measures have also evolved to include the assessment of a patient's suitability for surgery, expected outcome and post-surgical recovery [(4, 9-23)].

Since PROMs inception in the 1970s, their use has increased in parallel with the advent of evidence-based medicinal practice[(4)]. PROMs have evolved to include the assessment of patients' suitability for surgery; the expected outcome and efficacy of surgical outcomes; and the monitoring of patients' functional change over time [(4-23)]. They are also valuable research tools that are being used with increasing frequency. In the clinical setting, PROMs have been used to guide surgical decision-making and enhance communication and understanding during doctor- the best reported PROM patient consultations [(3, 5-7, 14, 24, 25)].

A wide variety of PROMs have been described, investigated and validated [4-23, 26-39]. Each assessment tool has its relative strengths and benefits, as well as its weaknesses. No single assessment tool has demonstrated consistent superiority over another[6-8]. In general, the choice of which one to use comes down to patient population, the pathology being investigated, investigator preference and resource management [33, 35, 37-39]. Some PROMs combine subjective patient-derived inputs with objective, clinician-derived inputs to establish a single score [19, 21, 23]. Other assessment tools include subjective patient-derived reports only [10, 12, 14, 16-18, 23, 30, 33, 37].

Advances in joint arthroplasty have seen an increase in the survivability of implants and a decrease in cumulative revision rates over the past 30 years[40]. The long-term results of joint arthroplasty require longitudinal patient and implant assessments. Accurate interpretation of these longitudinal studies requires an examination of patient, surgical and implant factors that may change over time[40]. PROMs are often used to augment the data collected by registries, thus adding important clinical data to supplement the revision data collected by arthroplasty registries [41-46].

PROMs are now commonly used in the evaluation of hip[13, 23], knee [10, 18] and shoulder [12, 14, 15, 47-50] arthroplasty patient cohorts. As these cohorts often include a higher proportion of elderly participants, an accurate interpretation of clinical scoring data relies on an understanding that variation may exist in age, ethnicity, nationality and gender [40, 51].

When considering PROM reports for pre-operative and post-operative clinical measures, these are best interpreted in comparison to asymptomatic, healthy, pathology-free, age- and gender-matched individuals[52]. Indeed, a perfect PROM score may not reflect a realistic post-operative goal. Several studies have confirmed that in an asymptomatic population, the best reported PROM scores were not equivalent to a perfect score on the outcome scale used [48, 53, 54]. In fact, an accurate interpretation of a patient's PROM score requires a comparison with an age and gender-matched group of individuals who have not had recent surgery or joint arthroplasty [48, 54-56].

Minimal research has been done to compare "normal" PROM values from different countries [57, 58]. Recent technological advances have enabled electronic data-collection. This presents the opportunity to collect uniform data, quickly, efficiently and accurately across continents [58, 59]. The collation of these data (with adherence to local government privacy requirements) enables the opportunity to establish a robust data collection system for future use when assessing pathological patient cohorts [60].

Variation in clinical PROM scoring is known to exist in different countries [25, 57]. This can make research outcomes derived in one geographical or cultural region difficult to generalize to regions of different ethnicity and socioeconomic diversity.

The establishment of a de-identified, asymptomatic "normal" control group within an electronic database would aid in improving the quality of future studies[61]. An international database has the capacity to include de-identified data from different regions remotely. These data have the potential to make future multi-centre, international clinical studies more robust and capable of rapidly advancing our understanding of clinical outcomes. Overall, this has the potential to significantly improve the ability to collect reliable, validated data and enhance our practice of evidence-based medicine.

The studies presented in this thesis represent the collection of independent asymptomatic participant cohorts for the hip, knee, shoulder and wrist, using commonly used PROM scores for each joint under investigation. The PROMs chosen represent internationally recognised, validated and commonly used clinical PROM scoring systems [10, 12-15, 18, 23, 47-50]. The rationale of each selected hip [13, 23], knee [10, 18], shoulder [12, 14, 15, 47-50] and wrist/hand PROM [11, 12], is discussed in the publications presented in this thesis. Ultimately, it is envisaged that the collection of independent, international asymptomatic PROM cohorts will aid in the establishment of control groups for comparison, for the evaluation of future longitudinal pathological patient cohorts.

The main aims of this thesis are:

1. To assess whether a variety of the most commonly used hip, knee, shoulder and wrist PROM scores are equivalent in asymptomatic, healthy, pathology-free individuals of different age, gender and ethnicity across two remote continents.

2. To establish normal population PROM values, using an electronic data collection system for future comparisons with pathological patient cohorts.

The hypothesis is that there is no difference in PROM clinical score values in an asymptomatic population when comparing age, gender, handedness (where applicable), history of pathology, ethnicity and nationality. If no difference is found to exist, then any PROM could be interpreted at face value, with a score of 100% being assumed as the goal for all post-operative patients, irrespective of these variables.

Each Chapter presented in this thesis represents novel research undertaken during candidature and is divided by the separate joint PROMs under investigation. Chapters 2 though to 5, represent the

research undertaken in reference to the hip, knee, shoulder and wrist/hand cohorts, respectively. Each cohort represent separate datasets collected during candidature, which are collated within the same electronic database and are de-identified. It is anticipated that these de-identified electronic datasets will be stored and used as control group(s) to aid in improving the quality of future studies.

CHAPTER 2 - NORMAL POPULATION REFERENCE VALUES FOR THE OXFORD & HARRIS HIP SCORES – ELECTONIC DATA COLLECTION AND ITS IMPLICATIONS

As published in Hip International. 2017; 27 (4): 389-396.

Introduction

Hip Patient Recorded Outcome Measures (PROMs) are now commonly used to evaluate the efficacy of hip arthroplasty patients both before and following surgery. The Harris Hip Score (HHS) [23] and Oxford Hip Score (OHS) [13] are the most widely used. Both measures have been shown to have acceptable reliability and construct validity. Accurate interpretation of long-term studies requires an understanding of not only the surgical and implant factors, but also patient factors prior to surgery and as they change over time.

Aims

The aim of this study was to assess whether the HHS and OHS were comparable in normal, healthy, pathology-free individuals of different ages, gender, ethnicity and nationality. The purpose of this study was to establish normal population values for the HHS and OHS, using an electronic data collection system. The hypothesis was that there is no difference in the HHS and OHS values for the asymptomatic "normal" population when comparing groups that differ in age, gender, ethnicity and nationality.

Main Findings

This study represents the largest database of asymptomatic "normal" HHS and OHS PROM values reported in the literature (n=627). No other study has compared normal scores in international cohorts or collected them remotely and electronically via the same research database. Other researchers have reported asymptomatic values for other musculoskeletal assessment tools, but few have approached the numbers collected in this study; with most collecting less than 150 participants or the participants were limited to young, active individuals.

There was an association between the OHS and age and also the HHS and age; demonstrating that as age increased, asymptomatic "normal" hip scores decreased. In comparing countries, the Australian asymptomatic group reported a statistically significant higher OHS score compared to Canadians. Overall, this study suggests that asymptomatic "normal" hip PROM values are comparable for individuals under the age of 80 years and are highly variable in individuals over the age of 80 years, making comparisons in this older age group less reliable. This finding suggests that there are likely

many other determinants of health and function that may influence the subjective hip PROM score reported by asymptomatic individuals and possibly the accuracy of asymptomatic "normal" values in this age group.

This study found that objective assessments contributed less than a 1/100 point difference in 12 % of participants and less than a 4/100 points difference in 0.5% of the study population. This finding suggests that resource management and cost justification should be considered when choosing a hip PROM tool for future studies, as its inclusion affected less than 13% of the study population in this asymptomatic "normal" cohort. A subjective-only hip PROM assessment tool has several advantages, including remote administration, thereby negating the need be reviewed by a clinician and saving resources. Using a subjective-only hip PROM would likely increase the cost-effectiveness of following surgical outcomes, while also minimising valuable resources and clinician time.

Future Directions

This study has established an electronic, "normal" control group for studies using the HHS and OHS PROMs. Studies which include such an electronic control group should consider differences in gender, age, ethnicity and nationality when using hip PROMs to assess patient outcomes. An accurate interpretation of the HHS and OHS score in a patient with hip pathology requires a comparison with an age- and gender-matched groups of individuals who have not had hip arthroplasty surgery.

This study found that when using the OHS, the control group should be sourced from the same country of origin. On the other hand, when using the HHS, the control group can be sourced from a pre-established control group within the database, without necessarily being sourced from the same country of origin. This study indicates that using a hip PROM, which does not include an objective component, can be considered a reliable and valid measure of patient outcome. Further research is required to determine if this principle is applicable to a pathological cohort. The use of a subjective-only hip PROM is particularly relevant when cost-justification and resource management are important.

Statement of Authorship & Contribution

This section provides a statement of the contribution of each author for all the peer review publications in the thesis. All co-authors have signed the statement of contribution.

Signed statements follow.

Statement of Authorship

Title of Paper	Normal population reference values for the Oxford and Harris Hip Scores – electronic data collection and its implications for clinical practice
Publication Status	☐ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted w ork w ritten in manuscript style
Publication Details	Hip International, 2017; 27 (4): 389-396.

Principal Author

Name of Principal Author (Candidate)	Dr James Marcus McLean		
Contribution to the Paper	Conceptual project design Etthics committee applications (Canada & Australia) Data collection (oversight & manual collection) Data collation & processing Manuscript preparation & submission Editor responses & preparing of re-submissions		
Overall percentage (%)	80%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	23 May 2018

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Dr Jacob Cappelletto		
Contribution to the Paper	Data collection (Australia) Pre-liminary data collation, processing & analyses Presentation of preliminary data		
Signature		Date	02 June 2018

Name of Co-Author	Dr Jock Clarnette		
Contribution to the Paper	Data collection (Australia) Pre-liminary data collation, processing & analyses Presentation of preliminary data		
Signature		Date	31 May 2018
Name of Co-Author	Dr Catherine Hill		
Contribution to the Paper	Conceptual project design consultant Pre-liminary data analyses Manuscript preparation Editor responses		
Signature		Date	30 May 2018

Name of Co-Author	Dr Tiffany Gill		
Contribution to the Paper	Conceptual project design consultant Pre-liminary data analyses Manuscript preparation Editor responses		
Signature		Date	31 May 2018

Name of Co-Author	Dr Daniel Mandziak		
Contribution to the Paper	Project oversight (Australia) Pre-liminary data collation, processing & analyses Manuscript preparation Editor responses		
Signature	-	Date	31 May 2018

Name of Co-Author	Dr Jordan Leith		
Contribution to the Paper	Project oversight (Canada) Etthics committee application (Canada) Manuscript preparation		
Signature		Date	29 May 2018

Published Manuscript

NORMAL POPULATION REFERENCE VALUES FOR THE OXFORD & HARRIS HIP SCORES – ELECTONIC DATA COLLECTION AND ITS IMPLICATIONS

Hip International. 2017; 27 (4): 389-396.

HIP
ISSN 1120-7000

Hip Int 2017; 27 (4): 389-396
DOI: 10.5301/hipint.5000465

ORIGINAL RESEARCH ARTICLE

Open Access

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 Cappelletto[1], Jock Clarnette[1], Catherine L. Hill[3], Tiffany Gill[3], Daniel Mandziak[1], Jordan Leith[2]

[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)

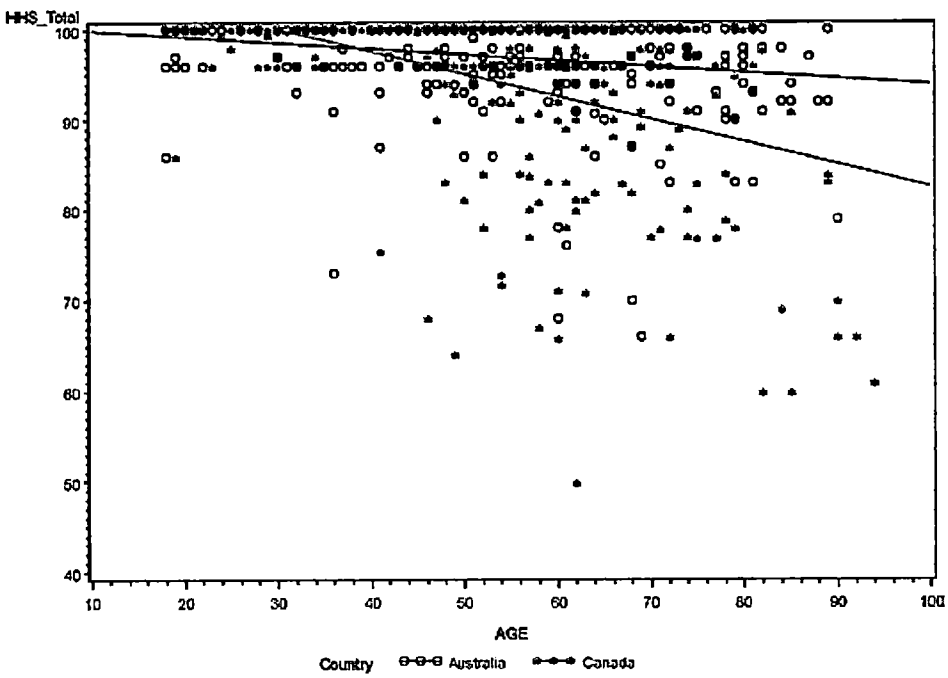

Fig. 1 - A scattergram of the Harris Hip Score versus age. Maximum score = 100 points. Red - Canada; Black - Australia.

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

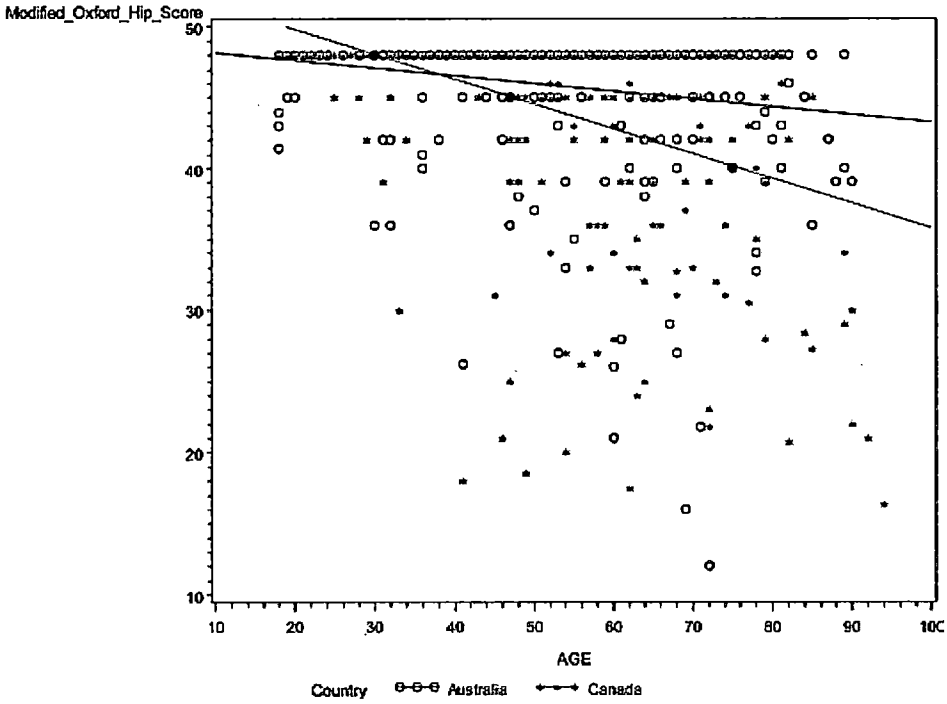

Fig. 2 - A scattergram of the Oxford Hip Score versus age. Maximum score = 48 points. Red - Canada; Black - Australia.

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, Tijssen 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, Bojescul 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.

CHAPTER 3 - ASYMPTOMATIC POPULATION REFERENCE VALUES FOR THREE KNEE PATIENT-REPORTED OUTCOME MEASURES – EVALUATION OF AN ELECTRONIC DATA COLLECTION SYSTEM & IMPLICATIONS FOR FUTURE INTERNATIONAL, MULTI-CENTRE COHORT STUDIES

As published in Archives of Orthopaedics & Trauma Surgery. 2018;138(5): 611-621.

Introduction

Knee Patient Recorded Outcome Measures (PROMs) are now commonly used to evaluate the efficacy of anterior cruciate ligament reconstruction and knee arthroplasty surgery on a longitudinal basis. The Knee Society Score (KSS)[18], Oxford Knee Score (OKS) [10] and Knee Injury and Osteoarthritis Outcome Score (KOOS) [29], are the most widely used. All three measures have been shown to have acceptable reliability and construct validity [27, 30, 31, 62]. Accurate interpretation of long term studies requires an understanding of not only the surgical and implant factors, but also patient factors, both before surgery and as they change over time.

Aims

The aim of this study was to assess whether the KSS, OKS and KOOS were comparable in normal, healthy, pathology-free individuals of different age, gender, ethnicity and nationality. The purpose of this study was to establish normal population values for the KSS, OKS and KOOS using an electronic data collection system. The hypothesis was that there is no difference in the KSS, OKS and KOOS scores in the asymptomatic "normal" population in regard to age, gender, ethnicity and different nationality.

Main Findings

This study represents the largest database of asymptomatic "normal" KSS, OKS and KOOS PROM values reported in the literature (n=614). No other study has compared normal scores in international cohorts or collected them remotely and electronically via the same research database. Other researchers have reported asymptomatic values for some of the PROMs examined in this study, but few have approached the numbers collected in this study [63, 64]; or the participants were limited to young, active individuals [65, 66].

There was an inverse relationship between the KSS and age, OKS and age; and KOOS and age. The specific finding was that as age increased, reported normal knee scores decreased. This study also demonstrated comparable PROM scores between national cohorts and between genders. A high percentage of participants (27%) reported *a history of an inactive knee problem,* which was associated with 5% lower reported normal knee scores when compared to those who had no such history. Participants who had experienced *a history of an active hip, ankle or foot problem* reported 10% lower

normal knee scores compared to those who had no such history. These findings suggest that there may well be an overlap in the functional questions contained in the assessment tools. This suggests that these three functional tools (KSS, OKS and KOOS) may not represent knee-specific PROMs. Consequently, these three measures may fail to discriminate between a primarily knee disability and other sources of lower limb incapacity.

Future directions

This study has established an electronic, "normal" control group for studies using the KSS, OKS and KOOS PROMs. When using these PROMs the control group can be drawn from a pre-established control group within an existing database, without necessarily being sourced from the same country of origin. Care should be taken when using the KSS, OKS and KOOS PROMs in patients with multiple, concurrent lower limb pathology, or in patients with a history of an inactive knee problem, as these factors may well have an impact the PROM values reported.

Studies using an electronic control group should consider differences in gender, age, ethnicity and nationality when using knee PROMs to assess patient outcomes. An accurate interpretation of the KSS, OKS and KOOS score in patients with knee pathology requires a comparison with an age- and gender-matched groups of individuals who have not had recent knee surgery or knee arthroplasty surgery. Consideration should be given to the influence of other lower limb pathology and contralateral disease on the three PROMs used in this study when assessing a pathological cohort.

Statement of Authorship & Contribution

This section provides a statement of the contribution of each author for all the peer review publications in the thesis. All co-authors have signed the statement of contribution.

Signed statements follow.

Statement of Authorship

Title of Paper	**Asymptomatic Population Reference Values for Three Knee Patient-Reported Outcomes Measures – Evaluation of an Electronic Data Collection System and Implications For Future International, Multi-Centre Cohort Studies**
Publication Status	☐ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted w ork w ritten in manuscript style
Publication Details	Archives of Orthopaedics & Trauma Surgery. 2018 May;138(5): 611-621

Principal Author

Name of Principal Author (Candidate)	Dr James Marcus McLean		
Contribution to the Paper	Conceptual project design Ethics committee applications (Canada & Australia) Data collection (oversight & manual collection) Data collation & processing Manuscript preparation & submission Editor responses & preparing of re-submissions		
Overall percentage (%)	80%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	23 May 2018

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Dr Oscar Brumby-Rendell		
Contribution to the Paper	Preliminary data collation, processing & analyses Presentation of preliminary data Manuscript preparation & Editor responses		
Signature		Date	30 May 2018

Name of Co-Author	Dr Ryan Lisle		
Contribution to the Paper	Data collection (Canada) Preliminary data collation, processing & analyses Manuscript preparation		
Signature		Date	10 Mar 2019

Name of Co-Author	Dr Jacob Brazier		
Contribution to the Paper	Data collection (Australia) Preliminary data collation, processing & analyses Presentation of preliminary data		
Signature		Date	12 July 2018

Name of Co-Author	Dr Kieran Dunn		
Contribution to the Paper	Data collection (Australia) Preliminary data collation, processing & analyses Presentation of preliminary data		
Signature		Date	12 Mar 2019

Signature		Date	
Name of Co-Author	Dr Tiffany Gill		
Contribution to the Paper	Conceptual project design consultant Preliminary data analyses Manuscript preparation Editor responses		
Signature		Date	31 May 2018

Name of Co-Author	Dr Catherine Hill		
Contribution to the Paper	Conceptual project design consultant Preliminary data analyses Manuscript preparation Editor responses		
Signature	Manuscript preparation	Date	30 May 2018

Name of Co-Author	Dr Daniel Mandziak		
Contribution to the Paper	Project oversight (Australia) Preliminary data collation, processing & analyses Manuscript preparation Editor responses		
Signature		Date	06 June 2018

Name of Co-Author	Dr Jordan Leith		
Contribution to the Paper	Project oversight (Canada) Ethics committee application (Canada) Manuscript preparation		
Signature		Date	29 May 2018

Published Manuscript

ASYMPTOMATIC POPULATION REFERENCE VALUES FOR THREE KNEE PATIENT-REPORTED OUTCOME MEASURES – EVALUATION OF AN ELECTRONIC DATA COLLECTION SYSTEM & IMPLICATIONS FOR FUTURE INTERNATIONAL, MULTI-CENTRE COHORT STUDIES

Archives of Orthopaedics & Trauma Surgery. 2018;138(5): 611-621.

Archives of Orthopaedic and Trauma Surgery
https://doi.org/10.1007/s00402-018-2874-4

ORTHOPAEDIC SURGERY

Asymptomatic population reference values for three knee patient-reported outcomes measures: evaluation of an electronic data collection system and implications for future international, multi-centre cohort studies

James M. McLean[1,2,4] · Oscar Brumby-Rendell[2] · Ryan Lisle[1] · Jacob Brazier[2] · Kieran Dunn[2] · Tiffany Gill[3] · Catherine L. Hill[3] · Daniel Mandziak[2] · Jordan Leith[1]

Received: 29 June 2017

Abstract

Objectives The aim was to assess whether the Knee Society Score, Oxford Knee Score (OKS) and Knee Injury and Osteoarthritis Outcome Score (KOOS) were comparable in asymptomatic, healthy, individuals of different age, gender and ethnicity, across two remote continents. The purpose of this study was to establish normal population values for these scores using an electronic data collection system.

Hypothesis There is no difference in clinical knee scores in an asymptomatic population when comparing age, gender and ethnicity, across two remote continents.

Methods 312 Australian and 314 Canadian citizens, aged 18–94 years, with no active knee pain, injury or pathology in the ipsilateral knee corresponding to their dominant arm, were evaluated. A knee examination was performed and participants completed an electronically administered questionnaire covering the subjective components of the knee scores. The cohorts were age- and gender-matched. Chi-square tests, Fisher's exact test and Poisson regression models were used where appropriate, to investigate the association between knee scores, age, gender, ethnicity and nationality.

Results There was a significant inverse relationship between age and all assessment tools. OKS recorded a significant difference between gender with females scoring on average 1% lower score. There was no significant difference between international cohorts when comparing all assessment tools.

Conclusions An electronic, multi-centre data collection system can be effectively utilized to assess remote international cohorts. Differences in gender, age, ethnicity and nationality should be taken into consideration when using knee scores to compare to pathological patient scores. This study has established an electronic, normal control group for future studies using the Knee society, Oxford, and KOOS knee scores.

Level of evidence Diagnostic Level II.

Keywords Assessment · Outcome · Score · Knee · Arthroplasty · Validation · Oxford · Electronic · Automated · Smartphone · iPhone · Laptop · Joint · Replacement

✉ James M. McLean
james.mc_lean@adelaide.edu.au; dr.jm.mclean@me.com

[1] Department of Orthopaedics, University of British Columbia, Vancouver, BC, Canada

[2] University of Adelaide Centre for Orthopaedic and Trauma Research, Royal Adelaide Hospital, North Terrace, Adelaide 5000, Australia

[3] Discipline of Medicine, Faculty of Health Sciences, University of Adelaide, Adelaide, SA, Australia

[4] Discipline of Orthopaedics and Trauma, University of Adelaide, North Terrace, Adelaide, SA 5005, Australia

Introduction

Patient-reported outcomes measures (PROMs) can be used to assess the efficacy of surgical outcomes, to monitor a patient's functional change over time, or as a research tool. More recently, knee PROMs have been used to guide surgical decision-making and enhance communication and understanding during doctor–patient consultations [1, 2].

PROM data has traditionally been collected in the clinic setting using a paper-based method. Newer, computer-based,

Published online: 12 January 2018

electronic PROM data collection systems allow for quicker data collection; automated data input and processing; and minimal clinician input [3].

Various knee PROMs have been described and validated [4–11]. Each assessment tool has its relative strengths and benefits, as well as its weaknesses. No single assessment tool has reported consistent superiority over another and the choice of which one to use normally comes down to patient population, the pathology being investigated, investigator preference and resource management.

The long-term results of common knee surgeries such as anterior cruciate ligament (ACL) reconstruction and total knee arthroplasty (TKA); require longitudinal patient and implant assessments. Accurate interpretation of these long-term studies requires an understanding of the patient, surgical and implant factors that may change over time.

When considering using knee PROMs for pre-operative and post-operative clinical scoring, these are best interpreted when compared to asymptomatic, healthy, pathology-free, age- and gender-matched individuals. An accurate interpretation of clinical scoring data relies on an understanding that variation may exist in ethnicity, nationality, gender and age. A perfect knee score may not reflect a realistic post-operative 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 recent surgery or a TKA.

The aim of this study was to assess whether the Knee Society Score (KSS), the Oxford Knee Score (OKS) and the Knee Injury and Osteoarthritis Outcome Score (KOOS), are equivalent in asymptomatic, healthy, pathology-free individuals of different age, gender and ethnicity, across two remote continents. The purpose of this study was to establish normal population values for the KSS, OKS and KOOS, using an electronic data collection system.

Our hypothesis was that there is no difference in knee clinical score values in an asymptomatic population when comparing age, gender, ethnicity and nationality.

Methods

Independent Ethics Board approval was granted from each Institution involved in the study. A power calculation determined the numbers required to reach statistical significance. From November 2014 to December 2015, healthy volunteers were recruited from a variety of sources. Participants were approached at Drivers Licensing Offices; public libraries; the Outpatients of both Public and Private Medical Facilities; and at various community centers (sporting, childcare, recreation and senior's activity facilities). There were no study advertisements and participants were not paid for their involvement.

Participants were included if they were 18 years of age or older; fluent in English; Australian or Canadian citizens; and reported both knees as "normal". For the purpose of the study where appropriate, participants were directed to answer the questions with reference to their ipsilateral knee corresponding to their dominant arm (ipsilateral knee). Exclusion criteria included participants with: a history of any inflammatory arthritis; significant lumbar spine problems that interfered with their mobility or function; or cognitive impairment or language problems. Participants were also excluded if they had active (i.e., painful or symptomatic) ipsilateral knee pathology; a history of ipsilateral knee joint arthroplasty; or ipsilateral knee surgery within the past 3 years. A history of inactive (i.e., no longer symptomatic) ipsilateral knee pathology, including previous surgery, was recorded, but was not considered part of the exclusion criteria. A history of active hip, ankle and/or foot pathology was recorded; as well as contralateral knee pathology (but these were also not considered part of the exclusion criteria). Eligible participants gave informed consent.

The study included 626 healthy volunteers, free of knee pathology. The Australian cohort included 162 (51.9%) privately insured participants with the remaining 150 (48.1%) being publically insured by the federal government. The Canadian federal government insures 100% of the Canadian population, representing an entirely publically funded health care system.

The Australian cohort included 155 males and 157 females. The average age was 53 years (range 18–93); 34 (10.9%) left and 278 (89.1%) right ipsilateral knees were included.

The Canadian cohort included 159 males and 155 females. The average age was 53 years (range: 18–91); 32 (10.2%) left and 282 (89.8%) right ipsilateral knees were included.

An electronic, web-based outcomes based electronic research database (OBERD, Universal Research Solutions, Columbia, Missouri, USA) was used to administer the KSS, OKS and the KOOS questionnaires. The details of these instruments are described in the following section.

Participants self-administered using an electronic mobile device (Smartphone, iPad or other electronic tablet device), or a laptop computer. If a participant had difficulty in completing the questionnaire, due to visual impairment, impaired dexterity, or computer unfamiliarity, a researcher completed it for them by verbally asking the questions and recording their responses. Where appropriate, participants were asked to complete the questions with reference to their ipsilateral knee corresponding to their dominant arm.

Participants were then assessed clinically and measurements of their pain-free, active range of motion (ROM), knee angulation and stability were recorded. ROM was assessed in the seated or supine position, using the axis of the femoral

and tibial bones, and the greater trochanter, distal femoral condyles and lateral malleolus as reference points. A single, highly experienced trained observer performed all of the assessments in North America. In Australia, two trained observers with less experience performed all of the assessments. Inter-variability correlation was performed.

The results were electronically transferred to the research database using OBERD. The combined subjective and objective assessments were combined electronically to determine the participant's KSS. This method enabled minimal data handling by the recruiters, ensuring that the investigators were partially blinded to the participants' results.

Primary outcome measures

Oxford Knee Score (OKS)

The OKS is a 12-item knee assessment tool. It is a purely subjective patient-report of pain, function and disability [2]. There are no objective, clinician-derived inputs. Each of the 12-items are scored from 0 to 4, with 4 representing best outcome/least symptoms. Scores from each question are added so the overall score ranges from 0 to 48, with 48 being the best score possible. The OSS score has been found to be simple and reliable, and to have excellent reliability and construct validity [4, 11].

Knee Society Score (KSS)

The KSS is a 7-item knee assessment tool. It is comprised primarily of objective data, with 6/7 of the inputs involving clinical assessments and the last input being a subjective, patient-report of pain. The clinician input includes an objective assessment of knee alignment, stability and ROM [5]. The responses are made on a visual analog scale, combined, and then scaled to a 100-point sum, where 100 is the best score possible [5]. The KSS has been shown to have acceptable reliability and construct validity [6, 10]. Scores are rated excellent (80–100); good (70–79); fair (60–69); or poor (less than 60).

Knee Injury and Osteoarthritis Outcome Score (KOOS)

The KOOS is a 42-item knee assessment tool. It is a purely subjective patient-report of knee symptoms, function and quality of life [8]. There are no objective, clinician-derived inputs. The responses are made on a visual analog scale, under the following sections: symptoms (KOOS-S: 5 for symptoms and 2 for stiffness); pain (KOOS-P: 9 items); function [17 items for daily activities (KOOS-DL) and 5 items for sport and recreational activities (KOOS-SR)]; and quality of life (KOOS-QOL: 4 items). The responses are combined, and then scaled to a 100-point sum, where 100 is the best score possible [8]. The KOOS has been shown to have acceptable reliability and construct validity [9].

Statistical analysis

Prior to analysis, the database was checked for missing data. Data analyses were performed using the IBM SPSS V.20 statistical package for Windows and SAS 9.3 (SAS Institute Inc., Cary, NC, USA).

A power calculation was performed by a statistician that determined the minimum sample size required to reach statistical significance ($n = 574$). Associations between nationality and age, gender, handedness and ethnicity, were investigated using chi-square tests and Fisher's exact test where appropriate. Due to the relatively low numbers recorded in some ethnic groups, no statistically significant comparisons could be made between the individual ethnic groups.

Poisson regression models were then used to investigate the association between knee scores, nationality, gender and age. Linear regression was not performed because, although the outcome variables (knee scores) were continuous, residuals from linear regression models were very left-skewed and could not be corrected using logarithmic transformation. The knee scores were therefore considered to be counts. Dispersion values for the Poisson regressions performed ranged from 0.0124 to 3.1994. Two-sided confidence intervals were set at 95% for two-way mixed effects model and absolute agreement. Interaction models were performed between nationality and age category; or gender versus the knee score outcomes. If the interaction was not significant, then main effects models were presented.

Results

There were no incomplete questionnaires. The cohort demographics are presented in Tables 1 and 2. There was no difference between the international cohorts when comparing age ($p = 0.7826$); gender ($p = 0.8107$); or handedness ($p = 0.7736$) (Table 1).

Overall, 27.2% Australians and 26.7% Canadians reported a history of an inactive knee problem, with 9.3% Australian and 9.2% Canadian of participants reporting having had non-arthroplasty knee surgery more than 3 years ago (Table 2). This difference was not statistically significant ($p = 0.9282$). There was a statistically significant association between all seven knee scores and participants with a history of an inactive knee problem (all seven models $p < 0.001$). For example, age- and gender-matched participants who reported a history of an inactive knee problem had a mean KSS score 4.8% lower than those who had no such history (IRR = 1.048; 95% CI: 1.029, 1.067). Similar statements could be made for the KOOS

Table 1 A comparison of ethnicity of the two international cohorts

Ethnicity	Australia ($n=312$)	Canada ($n=314$)	Total
Asian Indian	8 (2.5%)	13 (4.1%)	21 (3.4%)
Black or African American	3 (1%)	3 (1%)	6 (1%)
Caucasian	290 (93%)	233 (74%)	522 (83.4%)
Chinese	6 (1.9%)	29 (9.2%)	35 (5.6%)
Filipino	0	9 (2.9%)	9 (1.5%)
Indigenous	1 (0.3%)	2 (0.6%)	3 (0.5%)
Middle Eastern	2 (0.6%)	14 (4.5%)	16 (2.6%)
Other Asian	2 (0.6%)	8 (2.6%)	10 (1.6%)
Central/South American	0	3 (1%)	3 (0.5%)
	312	314	626 (100%)

Due to the relatively low numbers recorded in some ethnic groups, no statistically significant comparisons could be made between the individual ethnic groups

Table 2 Comparison of Australian and Canadian cohorts by demographics

Australian cohort ($n=312$)	Canadian cohort ($n=314$)	Total	Comparing Australian and Canadian cohorts
Male			
155 (49.7%)	159 (50.6%)	314 (50.2%)	$p=0.8107$
Female			
157 (50.3%)	155 (49.4%)	312 (49.8%)	
Left			
34 (10.9%)	32 (10.2%)	66 (10.5%)	$p=0.7736$
Right			
278 (89.1%)	282 (89.1%)	560 (89.5%)	
Age < 30			
30 (9.6%)	30 (9.6%)	60 (9.6%)	$p=0.9992$
Age 30–39			
35 (11.2%)	33 (10.5%)	68 (10.9%)	
Age 40–49			
53 (17.0%)	50 (15.9%)	103 (16.5%)	
Age 50–59			
73 (23.4%)	76 (24.2%)	149 (23.8%)	
Age 60–69			
70 (22.4%)	73 (23.3%)	143 (22.8%)	
Age 70–79			
35 (11.2%)	37 (11.8%)	72 (11.5%)	
Age 80+			
16 (5.1%)	15 (4.8%)	31 (5.0%)	
Average age			
53 (range 18–93)	53 (range 18–91)		
Privately insured			
162 (51.9%)	0		
Publically insured			
150 (48.1%)	314 (100%)		
Participant reported a history of an inactive (previous) knee problem			
85 (27.2%)	84 (26.7%)	169 (27%)	$p=0.9282$ See comment
Participant reported a history of an active hip, ankle or foot problem			
22 (7%)	39 (12.4%)	61 (9.7%)	$p=0.0253$
Participant reported having had non-arthroplasty knee surgery more than 3 years ago			
29 (9.3%)	29 (9.2%)	58 (9.3%)	$p=0.9283$

and OKS. There is a statistically significant association between report of history of inactive knee problem and age ($p = 0.0055$). This is an expected result, that as age increases the incidence of having had a knee problem increases.

A higher proportion of Canadian participants reported an active hip, ankle or foot problem (12.4 versus 7.0% of Australian participants; $p = 0.0253$; OR = 1.87, 95% CI:1.08, 3.23; Table 2). There was an association between all seven knee scores and participants with an active hip, ankle or foot problem (all seven models $p < 0.001$). For example, age- and gender-matched participants with an active hip, ankle or foot problem reported a mean OKS value 10.8% lower than those who had no such history (IRR = 1.108, 95% CI:1.064, 1.154). Similar statements could be made for the KOOS and OKS. Association between report of having an active hip, ankle of foot problem and age showed a trend, but did not reach statistical significance ($p = 0.0848$).

An inverse relationship was observed between age and PROM scores in all categories (Figs. 1, 2). PROM values are presented in Tables 3, 4, 5,6, and 7.

Oxford Knee Score

There was a statistically significant association between the OKS and age (Fig. 1), adjusting for nationality ($p < 0.0001$). For every 1-year increase in age, the mean OKS value decreased by 0.17% (Tables 3, 4; IRR = 0.9983; 95% CI: 0.9976, 0.9990).

There was no association between OKS and gender ($p = 0.7697$); handedness ($p = 0.6086$); or nationality, adjusting for gender ($p = 0.6837$) and adjusting for age ($p = 0.7406$).

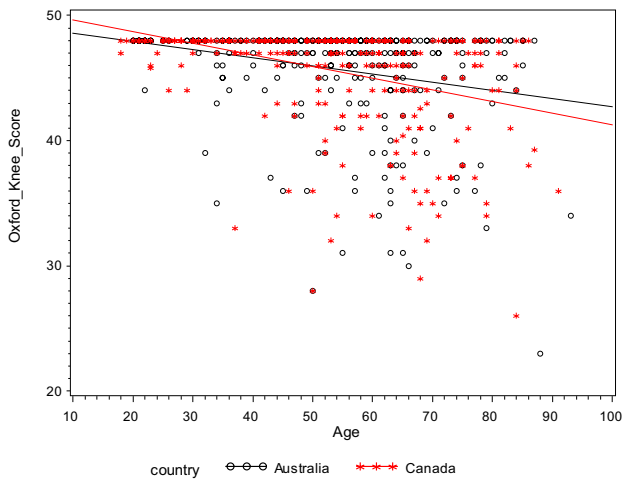

Fig. 1 Scatter plot of Oxford Knee Score versus age by country

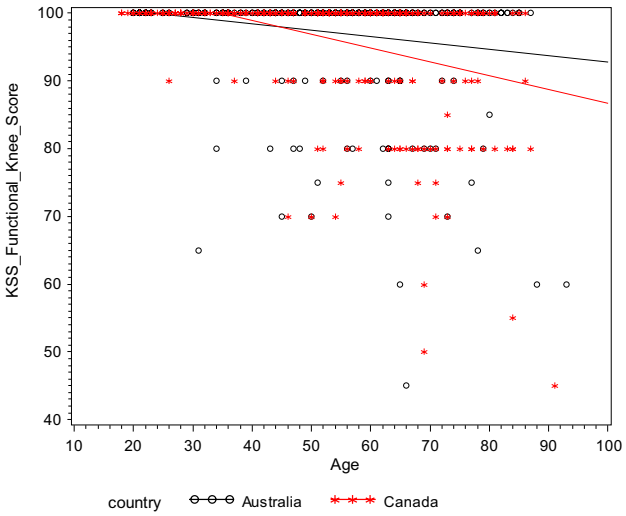

Fig. 2 Scatter plot of Knee Society Score versus age by country

Knee Society Score: clinician objective component

49 participants (7.8%) did not score the potential maximum 50/100 points for the clinician-assessed objective component. Of these participants, the average score was 1.2/50 (range 0.0–3.0). Forty one (6.6%) recorded a total ROM 70°–120° (mean total ROM = 110.0, range 75.2, 120.0; representing a loss of 1 to 10 points); 4 (0.6%) recorded an extension lag (allocated as a minus score of between −5 and −10); 25 (4.0%) reported a flexion contracture (allocated as a minus score of −2 to −15); 12 (1.9%) recorded an antero-posterior laxity greater than 5 mm (allocated as a minus score of −2 to −15); and 22 (3.5%) recorded medio-lateral laxity greater than 5 mm (representing a loss of 5–15 points).

Knee Society Score: total

There was a statistically significant association between OKS and KSS ($p < 0.0001$). For every one unit increase in OKS, the KSS value increased by 1.6% (IRR = 1.016, 95% CI: 1.014, 1.019).

There was a statistically significant association between the KSS and age (Fig. 2), adjusting for nationality ($p < 0.0001$). For every 1-year increase in age, the mean KSS value decreased by 0.15% (Tables 5, 6; IRR = 0.9985, 95% CI: 0.9980, 0.9990).

There was a statistically significant association between KSS and gender, adjusting for nationality ($p = 0.0124$). Specifically, female participants reported a mean KSS 1.0% lower than age-matched males (IRR = 0.9799; 95% CI: 0.9644, 0.9956).

There was no association between KSS and handedness ($p = 0.2837$); or KSS and nationality, adjusting for gender ($p = 0.1743$) and adjusting for age ($p = 0.2150$).

Table 3 Average Oxford Knee scores for Australian and Canadian cohorts, history of knee problem, current hip, ankle or foot and the combined cohorts

Age	Australia Ave	Canada Ave	*p* value	Hx of knee problem Ave	No Hx of knee problem Ave	*p* value	Current hip or ankle problem Ave	No current hip or ankle problem Ave	*p* value	Combined
< 30	47.8667	47.4606	0.8198	46.5000	47.7468	0.7272	44.0000	47.7257	0.5927	47.6636
30–39	46.6571	47.3636	0.7610	44.8000	47.6226	0.1593	46.0000	47.0462	0.7961	47.0000
40–49	46.5849	47.0800	0.7136	45.9600	47.1026	0.4675	45.0909	47.0326	0.3738	46.8252
50–59	46.0411	45.7620	0.8015	43.7674	46.7515	0.0150	40.0769	46.4552	0.0012	45.8987
60–69	44.3571	44.2590	0.9298	41.3400	46.0099	< 0.0001	39.8352	44.8704	0.0044	44.3071
70–79	44.7714	43.5135	0.4219	41.6000	45.4681	0.0187	41.8462	44.6271	0.1719	44.1250
80+	44.4375	43.0182	0.5505	36.8961	45.7500	0.0019	39.3182	44.4074	0.1512	43.7507
Overall	45.7756	45.5570	0.7327	42.9247	46.7003	0.2044	41.6006	46.1049	0.6597	45.666

Ave average, *combined* Australian and Canadian participants combined in that age group

Knee Injury and Osteoarthritis Outcome Score: total

There was no significant association between each KOOS subcategory and gender or handedness (Table 7).

There was a statistically significant association between each KOOS subcategory and age. For every 1-year increase in patient age, KOOS-DL decreased by 0.14% (IRR = 0.9986; 95% CI: 0.9981, 0.9991); KOOS-P decreased by 0.12% (IRR = 0.9988; 95% CI: 0.9983, 0.9993); KOOS-QOL decreased by 0.29% (IRR = 0.9971; 95% CI: 0.9966, 0.9976); KOOS-SR decreased by 0.45% (IRR = 0.9955; 95% CI: 0.9950, 0.9960); and KOOS-S decreased by 0.10% (IRR = 0.9990; 95% CI: 0.9985, 0.9995).

There was no association between KOOS subcategories and nationality, adjusting for gender and adjusting for age (Table 7).

Discussion

In this study, data were collected electronically from two asymptomatic, distinct, remote, Westernized populations of different countries, that were representative of their local populations [12, 13]. The ratio of participants of European descent (Caucasians by default) represented in the Australian cohort is consistent with that reported by the Australian Bureau of Statistics [12]. The higher proportions of Chinese, Middle Eastern and Asian Indians represented in the Canadian cohort, is consistent with that reported by the Government of Canada statistics [13].

To our knowledge, this study represents one of the largest databases of combined OKS, KSS and KOOS asymptomatic, normal values reported in the literature. Other studies have recorded normal values for knee outcome scores but few have approached the numbers collected in this study.

Paradowski collected normal reference values for KOOS on randomly chosen subjects in Sweden. Their study reported a response rate of 539/840, leading to response bias (as individuals with chronic pain issues are more likely to respond) [14]. Other studies have reported that subjects with a previous history of knee problems have a tendency to respond to medical surveys more readily than those without [15]. These studies were limited by comparing only a single PROM assessment tool; and limited to a single population cohort. These studies also used traditional paper assessments, requiring further data input for analysis.

The authors committed resources to collecting objective clinical data to complete the KSS. This study of asymptomatic individuals, reported that KSS, KOOS and OKS PROMs were comparable. This important finding led us to re-evaluate the importance of collecting objective data for calculating the KSS. As resource management and cost justification is becoming more of a focus for our Institutions,

Table 4 Confidence intervals and comparison *p* values for Oxford knee scores

Interaction	Group	Comparison	Incidence rate ratio	Lower 95% CL	Upper 95% CL	Comparison *p* value	Interaction *p* value
Country	< 30	Australia versus Canada	1.0086	0.9373	1.0853	0.8198	0.9704
	30–39	Australia versus Canada	0.9851	0.9191	1.0558	0.6710	
	40–49	Australia versus Canada	0.9895	0.9352	1.0470	0.7136	
	50–59	Australia versus Canada	1.0061	0.9595	1.0549	0.8015	
	60–69	Australia versus Canada	1.0022	0.9540	1.0528	0.9298	
	70–79	Australia versus Canada	1.0289	0.9598	1.1030	0.4219	
	80+	Australia versus Canada	1.0330	0.9286	1.1492	0.5505	
	Australia	80+ versus < 30	0.9284	0.8486	1.0157	0.1050	
	Canada	80+ versus < 30	0.9064	0.8259	0.9947	0.0384	
Hx of knee problem	< 30	Yes versus No	0.9739	0.8394	1.1299	0.7272	0.2044
	30–39	Yes versus No	0.9407	0.8640	1.0243	0.1593	
	40–49	Yes versus No	0.9757	0.9132	1.0426	0.4675	
	50–59	Yes versus No	0.9362	0.8877	0.9873	0.0150	
	60–69	Yes versus No	0.8985	0.8525	0.9470	< 0.0001	
	70–79	Yes versus No	0.9149	0.8496	0.9853	0.0187	
	80+	Yes versus No	0.8065	0.7042	0.9235	0.0019	
	Yes	80+ versus < 30	0.7935	0.6572	0.9580	0.0161	
	No	80+ versus < 30	0.9582	0.8932	1.0279	0.2333	
Hip or ankle problem	< 30	Yes versus No	0.9219	0.6845	1.2417	0.5927	0.6597
	30–39	Yes versus No	0.9778	0.8244	1.1596	0.7961	
	40–49	Yes versus No	0.9587	0.8737	1.0521	0.3738	
	50–59	Yes versus No	0.8627	0.7890	0.9433	0.0012	
	60–69	Yes versus No	0.8878	0.8180	0.9635	0.0044	
	70–79	Yes versus No	0.9377	0.8550	1.0284	0.1719	
	80+	Yes versus No	0.8854	0.7498	1.0455	0.1512	
	Yes	80+ versus < 30	0.8936	0.6397	1.2483	0.5095	
	No	80+ versus < 30	0.9305	0.8697	0.9955	0.0366	

consideration should be given to a knee PROM tool that is equivalent to the KSS, but does not require an objective assessment component. A subjective-only knee 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.

There was no association between KSS, KOOS and OKS and nationality, adjusted for age and gender. As the KSS, KOOS and OKS PROMs were comparable, this also suggests that future studies comparing pathological cohorts, can be performed with comparison against a pooled 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 electronic data collection, particularly the United States and Great Britain.

An electronically administered PROM enables more streamlined data collection, collating, processing and analyses. It is accessible to remote clinical researchers; allows simultaneous data entry and data analyses; has the potential to link to an electronic medical record; and enables patients the convenience of completing questionnaires remotely in their own time, or automating participant reminders or response time points. However, electronically administered questionnaires require respondents to be computer savvy; an assumption that may not be correct for all members of the public [16]. Patients with severe wrist and hand pathology may also be less willing to complete online questionnaires using computers or tablets.

Table 5 Average Knee Society Knee scores for Australian and Canadian cohorts, history of knee problem, current hip, ankle or foot and the combined cohorts

Age	Australia Ave	Canada Ave	*p* value	Hx of knee problem Ave	No Hx of knee problem Ave	*p* value	Current hip or ankle problem Ave	No current hip or ankle problem Ave	*p* value	Combined
< 30	100.00	99.6667	0.8972	100.00	99.8214	0.9725	100.00	99.8305	0.9866	99.8320
30–39	97.8571	99.6970	0.4454	98.0000	98.9623	0.7406	85.0000	99.3846	0.0143	98.7334
40–49	97.7358	98.8000	0.5861	98.0000	98.3333	0.8837	97.2727	98.3696	0.7287	98.2361
50–59	98.2877	96.6447	0.3098	95.3488	98.2857	0.1003	92.3077	97.9412	0.0494	97.4586
60–69	95.5714	94.8630	0.6643	90.7000	98.1522	< 0.0001	92.5000	95.5512	0.2385	95.2190
70–79	95.1429	91.6216	0.1222	93.2000	93.4043	0.9319	83.8462	95.4237	< 0.0001	93.3457
80+	94.0625	86.0000	0.0182	71.4286	95.6250	< 0.0001	77.5000	92.0370	0.0043	90.1448
Overall	97.1314	96.0828	0.1969	93.4024	97.8901	< 0.0001	90.2459	97.2920	0.0342	96.6054

Ave average, *Combined* Australian and Canadian participants combined in that age group

It is an important goal to differentiate normal, age-related changes in function, from those changes associated with pathological conditions. An inverse relationship was observed between age and PROM score, indicating that as age increases, the degree of baseline asymptomatic disability increases. There was a further significant inverse relationship within each subcategory of the KOOS, with the sports and recreations subcategory showing the largest decrease of 0.45% with each year of age. The greatest variation in PROM values were recorded in the ≥ 80 age group, with some respondents recording values of 48/48 and 100/100 (even up to 90 years); and others recording values half that of a perfect 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. This study also showed a significant association between age and reporting of inactive knee pathology confirming that as age increases a history or knee problems increases. There are likely many other determinants of health and function that may influence a participant's PROM score, and possibly the accuracy of "normal" values in this age group. Certainly, as patient age increases, perfect PROMs should not be expected following surgical interventions.

This study reported an association between a history of inactive knee pathology and all PROMs assessed. This seems logical and as such, should be taken into consideration when choosing an asymptomatic, normal control group to compare against a pathological or interventional group.

Participants with a history of dominant knee pathology treated by non-arthroplasty/arthrodesis surgery > 3 years ago and no longer experiencing symptoms were included. It is possible that some of these participants still suffered from some slight impairment as a result of their previous condition/surgery; however, we only included those who no longer perceived their issue to be a problem.

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

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 randomization. Although individuals were approached randomly in this study, no specific randomization method of participant identification was employed. Another potential source of interviewer bias involves the use of electronic questionnaires, where elderly

Table 6 Confidence intervals and comparison *p* values for Knee Society scores

Interaction	Group	Comparison	Incidence Rate Ratio	Lower 95% CL	Upper 95% CL	Comparison *p* value	Interaction *p* value
Country	< 30	Australia versus Canada	1.0033	0.9538	1.0555	0.8972	0.2089
	30–39	Australia versus Canada	0.9815	0.9357	1.0297	0.4454	
	40–49	Australia versus Canada	0.9892	0.9514	1.0285	0.5861	
	50–59	Australia versus Canada	1.0170	0.9844	1.0506	0.3098	
	60–69	Australia versus Canada	1.0075	0.9742	1.0419	0.6643	
	70–79	Australia versus Canada	1.0384	0.9899	1.0893	0.1222	
	80+	Australia versus Canada	1.0937	1.0154	1.1782	0.0182	
	Australia	80+ versus < 30	0.9406	0.8842	1.007	0.0526	
	Canada	80+ versus < 30	0.8629	0.8083	0.9211	< 0.0001	
Hx of knee problem	< 30	Yes versus No	1.0018	0.9051	1.1087	0.9725	< 0.0001
	30–39	Yes versus No	0.9903	0.9346	1.0492	0.7406	
	40–49	Yes versus No	0.9966	0.9523	1.0430	0.8837	
	50–59	Yes versus No	0.9701	0.9356	1.0059	0.1003	
	60–69	Yes versus No	0.9241	0.8917	0.9576	< 0.0001	
	70–79	Yes versus No	0.9978	0.9489	1.0492	0.9319	
	80+	Yes versus No	0.7470	0.6781	0.8228	< 0.0001	
	Yes	80+ versus < 30	0.7143	0.6263	0.8147	< 0.0001	
	No	80+ versus < 30	0.9580	0.9125	1.0057	0.0832	
Hip or ankle problem	< 30	Yes versus No	1.0017	0.8220	1.2206	0.9866	0.0342
	30–39	Yes versus No	0.8553	0.7547	0.9693	0.0143	
	40–49	Yes versus No	0.9888	0.9281	1.0535	0.7287	
	50–59	Yes versus No	0.9425	0.8884	0.9998	0.0494	
	60–69	Yes versus No	0.9681	0.9172	1.0217	0.2385	
	70–79	Yes versus No	0.8787	0.8235	0.9375	< 0.0001	
	80+	Yes versus No	0.8421	0.7483	0.9476	0.0043	
	Yes	80+ versus < 30	0.7750	0.6186	0.9709	0.0267	
	No	80+ versus < 30	0.9219	0.8797	0.9662	0.0007	

participants may have declined to be involved due to the technology.

Participants with a history of a prior knee injury may have chosen not to participate in the study, citing that their knee was not "normal". Although we chose to exclude participants with active knee pathology, we did include participants with a history of a previous knee 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 knee problem may have only minor functional incapacities, and should have been included in the study. As no X-rays were taken to confirm whether participants had asymptomatic degenerative knee disease, it is possible that some of these individuals were included in the cohorts. It is assumed that a small percentage of the general population will have asymptomatic or incidental undiagnosed knee pathology; however, we attempted to exclude subjects with active knee pathology as much as was practical. Previous studies reporting the collection of "normative values" have not made attempts to exclude subjects with pathology to the same extent this study has.

The carefully chosen inclusion and exclusion criteria was aimed at collecting an asymptomatic population who identified themselves as having 'normal' knees. Participants initially identified themselves as having no knee problems. Participants were then directed to answer the remainder of the questionnaire in reference to their knee that corresponded to their dominant hand. The Authors acknowledge that a participant's PROM value may be influenced by their contralateral knee, in a similar manner

Table 7 Average Knee Injury and Osteoarthritis Outcome Score (KOOS) for Australian and Canadian cohorts, and the combined cohorts

KOOS	Age	Gender	Handedness	Nationality
Daily living (KOOS-DL)	$p<0.0001$ (adjusting for Nationality)	$p=0.3464$ (adjusting for Nationality)	$p=0.3208$ (adjusting for Nationality)	$p=0.4216$ (adjusting for age) $p=0.3689$ (adjusting for gender)
Pain (KOOS-P)	$p<0.0001$ (adjusting for Nationality)	$p=0.9744$ (adjusting for Nationality)	$p=0.3737$ (adjusting for Nationality)	$p=0.8125$ (adjusting for age) $p=0.7543$ (adjusting for gender)
Quality of Life (KOOS-QOL)	$p<0.0001$ (adjusting for Nationality)	$p=0.5957$ (adjusting for Nationality)	$p=0.1320$ (adjusting for Nationality)	$p=0.6987$ (adjusting for age) $p=0.8277$ (adjusting for gender)
Sport and recreation (KOOS-SR)	$p<0.0001$ (adjusting for Nationality)	$p=0.9849$ (adjusting for Nationality)	$p=0.4364$ (adjusting for Nationality)	$p=0.4609$ (adjusting for age) $p=0.3116$ (adjusting for gender)
Symptoms (KOOS-S)	$p<0.0001$ (adjusting for Nationality)	$p=0.1560$ (adjusting for Nationality)	$p=0.3063$ (adjusting for Nationality)	$p=0.0769$ (adjusting for age) $p=0.0654$ (adjusting for gender)

Ave average, *Combined* Australian and Canadian participants combined in that age group. α adjusting for age. β adjusting for gender

to a pathological patient's PROM score being influenced by their unreported, contralateral knee function. The Authors believe this represents a variation in the pathological data sets. It is assumed that the contralateral knees in this study would have a similar distribution of coincident asymptomatic or undiagnosed pathology/normal contralateral knee function, as the pathological cohorts that it will be compared to in the future.

Concusion

Differences in age and gender should be taken into consideration when using the OKS, KSS and KOOS to assess patient outcomes. This study demonstrated comparable scores between national cohorts, suggesting that future international studies using these assessment tools can be performed using a standardized 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 using an electronic database.

Acknowledgements The authors would like to thank: Mami Okada, Research Assistant, University of British Columbia, for her help in the establishment of the study in Canada. Tara-Louise McLean, Research Assistant to Dr James McLean, for her help in the establishment of the study in Australia and Canada. Suzanne Edwards, Statistician, Data Management and Analysis Centre at the University of Adelaide, for her help with the statistical analysis of the collected data.

Funding 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.

Compliance with ethical standards

Conflict of interest All named authors, their immediate families, and any research foundations with which they are affiliated, have no conflicts of interest to disclose and no commercial associations (e.g., consultancies, stock ownership, equity interest, patent/licensing arrangements, etc) that might pose a conflict of interest in connection with the submitted article.

References

1. Coons SJ, Gwaltney CJ, Hays RD et al (2009) Recommendations on evidence needed to support measurement equivalence between electronic and paper-based patient-reported outcome (PRO) measures: ISPOR ePRO good research practices task force report. Value Health 12(4):419–429
2. Dawson J, Fitzpatrick R, Carr A (1996) Questionnaire on the perceptions of patients about shoulder surgery. J Bone Jt Surg Br 78(4):593–600
3. Griffiths-Jones W, Norton MR, Fern ED, Williams DH (2014) The equivalence of remote electronic and paper patient reported outcome (PRO) collection. J Arthroplasty 29(11):2136–2139
4. Dunbar MJ, Robertsson O, Ryd L, Lidgren L (2001) Appropriate questionnaires for knee arthroplasty. J Bone Jt Surg 83(3):339–344
5. Ewald FC (1989) The knee society total knee arthroplasty roentgenographic evaluation and scoring system. Clin Orthop Relat Res 248:9–12

6. Lingard EA, Katz JN, Wright RJ, Wright EA, Sledge CB (2001) Validity and responsiveness of the knee society clinical rating system in comparison with the SF-36 and WOMAC. J Bone Jt Surg Am 83(12):1856–1864
7. Murray DW, Fitzpatrick R, Rogers K et al (2007) The use of the Oxford hip and knee scores. J Bone Jt Surg Br 89-B(8):1010–1014
8. Roos EM, Roos HP, Lohmander LS, Ekdahl C, Beynnon BD (1998) Knee Injury and Osteoarthritis Outcome Score (KOOS)—development of a self-administered outcome measure. J Orthop Sports Phys Ther 28(2):88–96
9. Roos EM, Toksvig-Larsen S (2003) Knee injury and Osteoarthritis Outcome Score (KOOS)—validation and comparison to the WOMAC in total knee replacement. Health Qual Life Outcomes 1:17
10. Silva AL, Demange MK, Gobbi RG, da Silva TF, Pecora JR, Croci AT (2012) Translation and validation of the Knee Society Score—KSS for Brazilian Portuguese. Acta Ortop Bras 20(1):25–30
11. V.Harcourt WG, White SH, Jones P (2001) Specificity of the Oxford knee status questionnaire. J Bone Jt Surg 83(3):345–347
12. Cultural Diversity in Australia (2011) Reflecting a nation: stories from the 2011 Census. Australian Bureau of Statistics 2011, Canberra
13. Ethnic orgins for canada, provinces and territories Census of Population. G.o. Canada. Calgary. 2011
14. Paradowski PT, Bergman S, Sundén-Lundius A, Lohmander LS, Roos EM (2006) Knee complaints vary with age and gender in the adult population. Population-based reference data for the Knee injury and Osteoarthritis Outcome Score (KOOS). BMC Musculoskelet Disord 7:38
15. Baker P (2003) Knee disorders in the general population and their relation to occupation. Occup Environ Med 60(10):794–797
16. Snyder CF, Aaronson NK, Choucair AK et al (2011) Implementing patient-reported outcomes assessment in clinical practice: a review of the options and considerations. Qual Life Res 21(8):1305–1314

CHAPTER 4 - AN INTERNATIONAL, MULTI-CENTRE COHORT STUDY COMPARING SIX SHOULDER CLINICAL SCORES IN AN ASYMPTOMATIC POPULATION

As published in Journal of Shoulder & Elbow Surgery. 2018 Feb; 27(2): 306-314.

Introduction

Shoulder PROMs are commonly used to evaluate the efficacy of arthroscopic shoulder surgery, open rotator cuff surgery and shoulder arthroplasty surgery over a longitudinal period [25, 47, 67-71]. Several shoulder PROMs have been described and validated [14, 16, 17, 21, 47, 53, 56]. This thesis investigated six clinical measures including the American Shoulder and Elbow Surgeons (ASES) shoulder score [16]; the Constant-Murley Shoulder Score (CSS) [53]; the Oxford Shoulder Score (OSS) [14, 47]; the University of California, Los Angeles (UCLA) shoulder score [21]; the Shoulder Pain and Disability Index (SPADI) [17]; and the Stanmore Percentage of Normal Shoulder Assessment (SPONSA) [56]. Accurate interpretation of long-term studies requires an understanding of not only the surgical and implant factors, but also patient factors that may change over time.

Aims

The aim of this study was to assess whether the ASES, CSS, OSS, UCLA, SPADI, and SPONSA shoulder scores 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 six PROM shoulder scores being investigated, using an electronic data collection system. The hypothesis was that there is no difference in the PROM shoulder score values in an asymptomatic "normal" population when comparing sub-groups of differing age, gender, ethnicity and nationality.

Main Findings

This study represents the largest database of asymptomatic "normal" shoulder PROM values reported in the literature (n=635). No other study has compared six shoulder PROMs in an asymptomatic "normal" cohort. This study represents the only available data on normative SPADI [17] and SPONSA [56] values in the literature. Other researchers have reported asymptomatic values for some of the PROMs examined in this study [48, 54, 55, 72], but few samples have approached the numbers collected in this study. Frequently, the participants were limited to young, active individuals, which was not the case in the present study[48].

There was an association between the PROM shoulder scores being investigated and age, demonstrating that, as age increased, reported PROM scores were poorer. In individuals aged over 80 years old, there was a greater variation in PROM values reported, suggesting that comparisons within

this older age group may be less reliable. This finding suggests that there are likely many other determinants of health and function that may influence the subjective shoulder PROM score reported, and possibly the accuracy of asymptomatic "normal" values in this older age group.

Women reported similar or poorer PROM scores for all PROMs assessed, but poorer scores for females only reached statistical significance in three out of the six PROMs being investigated (ASES, CSS, and SPADI). Other studies have also reported poorer scores in asymptomatic "normal" females participants [53, 55].There was a statistically significant difference between the international cohorts for three out of the six PROMs (ASES, UCLA, and SPADI). This may reflect a true difference between the population cohorts, or a selection or reporting bias.

Fifteen percent of participants reported *a history of an inactive shoulder problem*. This sub-group reported 2% lower shoulder scores, compared to those who had no such history. This suggests that *a history of an inactive shoulder problem* should be taken into consideration when choosing an asymptomatic "control" group to compare against an interventional group.

Interestingly, participants who reported *a history of an active elbow, wrist or hand problem* reported similar PROM scores compared to those who had no such history. This finding supports the conclusion that these six shoulder PROMs are specific to the shoulder and can discriminate shoulder-based pathology from other sources of upper-limb incapacity. To the Author's knowledge, this finding has not previously been reported in the literature for these six PROMs.

There was a statistically significant correlation between all shoulder PROM assessment tools, with no difference identified when comparing subjective-only PROMs with combined subjective-objective PROMs. This finding suggests there may be no advantage in using a combined subjective / objective assessment tool, which requires additional resources, personnel and expertise to facilitate. To the Author's knowledge, this finding has not previously been reported in the literature for these six PROMs.

Future directions

An accurate interpretation of a shoulder PROM score from a patient with shoulder pathology requires a comparison with an age- and gender-matched group of individuals who have not had recent shoulder surgery or shoulder arthroplasty surgery. Studies using an electronic control group should consider differences in gender, age, ethnicity, handedness, history of pathology and nationality, when using shoulder PROMs to assess patient outcomes.

This study has established an electronic, asymptomatic “normal” control group for studies using the ASES, CSS, OSS, UCLA, SPADI, and SPONSA shoulder scores. When using the ASES, UCLA, and SPADI PROMs, the control group should be sourced from the same country of origin as the proposed study. Care should be taken when assessing patients aged over 80 years old, females, or those with a history of an inactive shoulder problem, as these factors are likely to have an impact on the PROM values reported.

Statement of Authorship & Contribution

This section provides a statement of the contribution of each author for all the peer review publications in the thesis. All co-authors have signed the statement of contribution.

Signed statements follow.

Statement of Authorship

Title of Paper	An International, Multi-Center Cohort Study Comparing Six Shoulder Clinical Scores in an Asymptomatic Population
Publication Status	☐ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted w ork w ritten in manuscript style
Publication Details	Journal of Shoulder & Elbow Surgery. 2018 Feb; 27(2): 306-314

Principal Author

Name of Principal Author (Candidate)	Dr James Marcus McLean		
Contribution to the Paper	Conceptual project design Etthics committee applications (Canada & Australia) Data collection (oversight & manual collection) Data collation & processing Manuscript preparation & submission Editor responses & preparing of re-submissions		
Overall percentage (%)	80%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	23 May 2018

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Dr Daniel Awwad		
Contribution to the Paper	Preliminary data collation, processing & analyses Presentation of preliminary data Manuscript preparation & Editor responses		
Signature		Date	30 May 2018

Name of Co-Author	Dr Ryan Lisle		
Contribution to the Paper	Data collection (Canada) Pre-liminary data collation, processing & analyses Manuscript preparation		
Signature		Date	10 Mar 2019

Name of Co-Author	Dr James Besanko		
Contribution to the Paper	Data collection (Australia) Pre-liminary data collation, processing & analyses Presentation of preliminary data		
Signature		Date	31 May 2018

Name of Co-Author	Dr Donald Shivakkumar		
Contribution to the Paper	Data collection (Australia) Pre-liminary data collation, processing & analyses Presentation of preliminary data		
Signature		Date	04 June 2018

Name of Co-Author	Dr Jordan Leith		
Contribution to the Paper	Project oversight (Canada) Etthics committee application (Canada) Manuscript preparation		
Signature		Date	29 May 2018

Published Manuscript

AN INTERNATIONAL, MULTI-CENTRE COHORT STUDY COMPARING SIX SHOULDER CLINICAL SCORES IN AN ASYMPTOMATIC POPULATION

Journal of Shoulder & Elbow Surgery. 2018 Feb; 27(2): 306-314.

J Shoulder Elbow Surg (2018) 27, 306–314

ELSEVIER

Journal of Shoulder and Elbow Surgery

www.elsevier.com/locate/ymse

An international, multicenter cohort study comparing 6 shoulder clinical scores in an asymptomatic population

James M. McLean, MS[a,b,*], Daniel Awwad, MBBS[a], Ryan Lisle, MBBS[b], James Besanko, MBBS[a], Donald Shivakkumar, MBBS[a], Jordan Leith, MD[b]

[a]*Discipline of Orthopaedics and Trauma, University of Adelaide, North Terrace, Adelaide, SA, Australia*
[b]*Department of Orthopaedics, University of British Columbia, Vancouver, BC, Canada*

Hypothesis: The study purpose was to assess 6 shoulder patient-reported outcome measure (PROM) values in asymptomatic, healthy, pathology-free individuals. We hypothesized that there would be no difference in PROM values in pathology-free individuals when considering sex, age, ethnicity, and geographical location.
Methods: Electronic questionnaires were completed by 635 individuals (323 Australians and 312 Canadians) without dominant shoulder pathology for the American Shoulder and Elbow Surgeons (ASES) shoulder score; Constant-Murley Shoulder Score (CSS); Oxford Shoulder Score (OSS); University of California, Los Angeles (UCLA) shoulder score; Shoulder Pain and Disability Index (SPADI); and Stanmore Percentage of Normal Shoulder Assessment (SPONSA). Shoulder range of motion and strength were assessed.
Results: No difference was identified between subjective-only and subjective-objective PROMs. Handedness and a current elbow or wrist problem were not associated with differences in PROM values. Poorer PROM values were associated with a history of an inactive shoulder problem and increasing age. Female participants tended to report similar or poorer PROM scores. No significant difference was found between ethnicities. Geographical location was associated with differences in the ASES shoulder score, UCLA shoulder score, and SPADI but not the CSS, SPONSA, and OSS.
Conclusions: Differences in sex, age, and geographical location will affect PROM shoulder scores in pathology-free individuals and should be taken into consideration when PROMs are being used to compare patient outcomes. This study has established normative values for the ASES shoulder score, CSS, OSS, UCLA shoulder score, SPADI, and SPONSA. Future studies assessing a pathologic patient cohort should perform comparisons against a sex- and age-matched control cohort, ideally sourced from the same geographical location.
Level of evidence: Basic Science Study; Validation of Outcome Instruments

Keywords: Assessment; outcome; shoulder score; validation; Constant; Oxford; SPADI

Independent ethics board approval was granted from each institution involved in the study: Royal Adelaide Hospital Human Research Ethics Committee (HREC reference No. HREC/14/RAH/325, RAH Protocol No. 140819) and University of British Columbia–Providence Health Care Research Institute (UBC-PHC REB No. H14-02030).

*Reprint requests: James M. McLean, MS, MBBS, FRACS, FAOrthA, Discipline of Orthopaedics and Trauma, University of Adelaide, North Terrace, Adelaide, SA 5005, Australia.
E-mail address: james.mc_lean@adelaide.edu.au (J.M. McLean).

1058-2746/$ - see front matter
https://doi.org/10.1016/j.jse.2017.08.016

The collection of preoperative and postoperative patient-reported outcome measures (PROMs) has traditionally been used to assess the efficacy of surgical interventions.[12] Since their inception, PROMs have been used with increasing frequency to measure the severity of a patient's symptoms and level of function. They can also be used as adjuncts to enhance communication and understanding during doctor-patient consultations.[7,9]

PROM data have traditionally been collected in the clinic setting using a paper-based method. Newer computer-based, electronic PROM data collection systems allow for remote data collection and questionnaire administration, automated data input and processing, quicker and real-time data collation, and minimal clinician input.[13]

Various shoulder PROM clinical scores have been described and validated. Several studies have confirmed that in an asymptomatic population, the best possible shoulder score may not be equivalent to a perfect score on the outcome scale used.[3,4] Preoperative and postoperative clinical scores are best interpreted when compared with normal, healthy, pathology-free age- and sex-matched individuals.[16] An accurate interpretation of clinical scoring data relies on an understanding that variation may exist with regard to sex, age, ethnicity, and geographical location.

The aim of this study was to assess whether 6 commonly used shoulder PROM clinical scores were equivalent in asymptomatic, healthy individuals of different sexes, ages, ethnicities, and geographical locations. The study compared subjective-only and combined subjective-objective PROMs with questionnaires administered and data collected electronically, including over 600 participants. The clinical scores under investigation included the American Shoulder and Elbow Surgeons (ASES) shoulder score; the Constant-Murley Shoulder Score (CSS); the Oxford Shoulder Score (OSS); the University of California, Los Angeles (UCLA) shoulder score; the Shoulder Pain and Disability Index (SPADI); and the Stanmore Percentage of Normal Shoulder Assessment (SPONSA).

Our hypothesis was that there would be no difference in shoulder PROM clinical scores in an asymptomatic population between sexes, age groups, ethnic groups, and geographical locations. If no difference existed, then any shoulder PROM could be interpreted at face value, with a score of 100% being assumed as the goal for all postoperative patients, irrespective of sex, age, ethnicity, or geographical location.

Methods

From November 2014 to November 2015, healthy volunteers were recruited from a variety of sources. Participants were approached at driver's licensing offices, public libraries, the outpatient services of both public and private medical facilities, and various community centers (sporting, childcare, recreation, and senior activity facilities). There were no study advertisements, and participants were not paid for their involvement.

Participants were included if they were aged at least 18 years; were fluent in English; were Australian or Canadian citizens; and had no diagnosed shoulder pathology in the dominant arm. The exclusion criteria included participants with a history of any inflammatory arthritis, significant neck problems, or cognitive impairment or language problems. Participants were also excluded if they had active dominant shoulder pathology or had a history of dominant shoulder surgery that included recent surgery (within the past 3 years) or joint arthroplasty. A history of inactive dominant shoulder pathology including previous surgery (>3 years ago) was recorded but was not considered part of the exclusion criteria. A history of ipsilateral elbow, wrist, or hand pathology was recorded, but this was also not considered part of the exclusion criteria.

Eligible participants underwent an informed consent process. The study included 635 participants free of active shoulder pathology (323 Australian and 312 Canadian citizens). The Australian cohort included 163 male and 160 female participants; the average age was 53.5 years (range, 20-89 years); 31 were left hand dominant. The Canadian cohort included 153 male and 159 female participants; the average age was 53.8 years (range, 19-90 years); 26 were left hand dominant.

An electronic, Web-based software system (OBERD [Outcomes Based Electronic Research Database]; Universal Research Solutions, Columbia, MO, USA) was used to combine several of the shoulder PROM instruments to create 1 condensed instrument. Condensed, electronically administered questionnaires have been shown to have comparable results to individually administered paper-based PROMs.[13,25] All questions were stated exactly as in the original instruments, added sequentially together. When appropriate, both imperial and metric values were stated. The shoulder outcome instruments assessed included the ASES shoulder score, CSS, OSS, UCLA shoulder score, SPADI, and SPONSA. The details of these instruments are described in the following section.

The questionnaire was self-administered by participants with reference to their dominant shoulder, using an electronic mobile device (smartphone or tablet computer) or a laptop computer. If participants had difficulty in completing the questionnaire because of computer unfamiliarity, visual impairment, or impaired dexterity, an investigator completed it for them by verbally asking the questions and recording their responses.

All participants were then assessed clinically, and measurements of their range of motion (ROM) and strength were recorded. ROM was assessed using the smartphone application DrGoniometer (version 1.9; CDM SrL, Milan, Italy).[20] Participants' pain-free active ROM was assessed in the seated position, using the axis of the arm and the spinous processes of the thoracic spine as reference points.[6] Subjective shoulder strength was assessed using the 6 grades (0-5) described by the Medical Research Council.[15] Objective shoulder strength was measured using an IDO Isometer (Innovative Design Orthopaedics, Reading, UK) and using the technique described by Constant et al.[6]

Primary outcome measures

ASES shoulder score

The ASES shoulder score is a 17-item patient report of pain, function, and disability, scored out of 100, which has been shown to have acceptable reliability and construct validity.[17,22]

Constant-Murley Shoulder Score

The CCS is an 8-item combined patient-clinician report of pain, function, ROM, and strength, which is scaled out of 100 points.[6] It has been found to be easy and quick to use and to allow interpretations comparable to those of other, more sophisticated scoring systems.[5,6]

Oxford Shoulder Score

The OSS is a 12-item patient report of pain, function, and disability with a maximal 48 points.[9] It is simple and reliable and has been shown to have excellent reliability and construct validity.[8,10]

UCLA shoulder score

The UCLA shoulder score is a 5-item combined patient-clinician report of pain, function, ROM, strength, and patient satisfaction. It is easy and quick to use and allows interpretations comparable to those of other, more sophisticated scoring systems.[19] This item is usually scored out of 35 points and was modified in our study to exclude points for postoperative patient satisfaction, giving a maximum score of 30 points.

Shoulder Pain and Disability Index

The SPADI is a 13-item patient report of shoulder pain and disability, with 0 as the best score and 100 as the poorest; it has been shown to have excellent reliability and construct validity.[18,23]

Stanmore Percentage of Normal Shoulder Assessment

The SPONSA is a single-question patient report with a score from 0 to 100 for shoulder pain, range of movement, strength, stability, and function. It has been found to be practical and quick to administer, with excellent reliability and construct validity.[21]

Statistical analysis

A power calculation was performed to determine the sample size necessary to detect a clinically significant difference in shoulder score of 20% at a power of 80% and an α value of .05 (n = 596). Associations between sex, age, ethnicity, and geographical location were investigated using Poisson and log-linear regression analysis. The primary outcome measures were not normally distributed on linear regression; hence, this analysis was not used. The SPADI outcomes were analyzed using log-linear regression as the results were normally distributed. For the remaining outcome measures (ASES shoulder score, CSS, OSS, UCLA shoulder score, and SPONSA), Poisson regression was used. These models fit the data well as the ratio of the Deviance to degrees of freedom, Value divided by degrees of freedom, was about 1. Estimates and 95% confidence intervals were exponentiated to give an incidence rate ratio and corresponding 95% confidence interval. Reliability studies were performed using Stata Statistical Software (release 14; StataCorp, College Station, TX, USA). All other analyses were performed with SAS software (version 9.4; SAS Institute, Cary, NC, USA).

Results

The cohort demographic data are presented in Tables S1 and S2. Because of the relatively low numbers recorded in some ethnic groups, no statistically significant comparisons could be made between the individual ethnic groups**.** When comparing white participants versus participants of other ethnicities, we found no significant differences (Table I).

Table I Comparison of shoulder scores in white participants and participants of other ethnicities

	Mean score		
	White participants	Participants of other ethnicities	*P* value
ASES	96	96	.7077
CSS	87	88	.2681
OSS	47	47	.8457
UCLA	29	29	.4506
SPADI	3.7	4.3	.6749
SPONSA	95	95	.8408

ASES, American Shoulder and Elbow Surgeons; *CSS*, Constant-Murley Shoulder Score; *OSS*, Oxford Shoulder Score; *UCLA*, University of California, Los Angeles; *SPADI*, Shoulder Pain and Disability Index; *SPONSA*, Stanmore Percentage of Normal Shoulder Assessment.

A history of an inactive shoulder problem was reported in 14.6% of Australians and 17% of Canadians (*P* = .16). Participants with a history of an inactive shoulder problem had poorer scores than those with no such history (Table II). This was statistically significant for all scores except the CSS. A current elbow or wrist problem (Table III) and handedness (Table IV) were not associated with differences in any scores.

When we compared sexes, women reported similar or poorer PROM scores for all PROMs assessed. However, when we controlled for geographical location and age, only 3 were statistically significant (ASES, CSS, and SPADI; Table V).

As age increased, PROM scores were poorer in all categories (Figs. 1 and S1).

PROM scores between geographical locations are presented in Figures 2-7 and Tables S3-S8. There were no statistically significant differences in PROM scores reported

Table II Comparison of shoulder scores in participants with history of shoulder problem more than 3 years ago that has resolved and participants without history

	Mean score		
	No history of shoulder problem	History of shoulder problem	Wilcoxon rank sum test (*P* value)
ASES	96.7	93.9	.0003
CSS	87.2	85.6	.1521
OSS	46.8	46.1	.0013
UCLA	28.9	28.3	.0005
SPADI	3.3	6.1	<.0001
SPONSA	94.3	92.3	.0116

ASES, American Shoulder and Elbow Surgeons; *CSS*, Constant-Murley Shoulder Score; *OSS*, Oxford Shoulder Score; *UCLA*, University of California, Los Angeles; *SPADI*, Shoulder Pain and Disability Index; *SPONSA*, Stanmore Percentage of Normal Shoulder Assessment.

Table III Comparison of shoulder scores in participants with and without current elbow or wrist problem

	Mean score		
	No current elbow or wrist problem	Current elbow or wrist problem	Wilcoxon rank sum test (*P* value)
ASES	96.3	92.6	.8159
CSS	86.9	90.1	.3369
OSS	46.7	46.4	.5870
UCLA	28.8	28.4	.2578
SPADI	3.8	3.0	.9169
SPONSA	93.9	96.1	.5920

ASES, American Shoulder and Elbow Surgeons; *CSS*, Constant-Murley Shoulder Score; *OSS*, Oxford Shoulder Score; *UCLA*, University of California, Los Angeles; *SPADI*, Shoulder Pain and Disability Index; *SPONSA*, Stanmore Percentage of Normal Shoulder Assessment.

Table IV Comparison of shoulder scores by handedness

	Mean score		
	Right handed	Left handed	*P* value
ASES	96	98	.2490
CSS	87	88	.4088
OSS	47	47	.5378
UCLA	29	29	.6150
SPADI	3.6	3.2	.7388
SPONSA	94	96	.1574

ASES, American Shoulder and Elbow Surgeons; *CSS*, Constant-Murley Shoulder Score; *OSS*, Oxford Shoulder Score; *UCLA*, University of California, Los Angeles; *SPADI*, Shoulder Pain and Disability Index; *SPONSA*, Stanmore Percentage of Normal Shoulder Assessment.

Table V Comparison of shoulder scores by sex adjusting for nationality

	Mean score		
	Men	Women	*P* value
ASES	97	95	.0303
CSS	90	84	<.0001
OSS	47	47	.4903
UCLA	29	28	.4768
SPADI	2.7	4.2	.0262
SPONSA	94	94	.7218

ASES, American Shoulder and Elbow Surgeons; *CSS*, Constant-Murley Shoulder Score; *OSS*, Oxford Shoulder Score; *UCLA*, University of California, Los Angeles; *SPADI*, Shoulder Pain and Disability Index; *SPONSA*, Stanmore Percentage of Normal Shoulder Assessment.

between Canadians and Australians for the CSS and SPONSA (Tables S4 and S8, respectively), as well as for the OSS in all age brackets younger than 80 years (Table S5). A statistically significant difference was reported between Canadians and Australians for ASES, UCLA, and SPADI scores

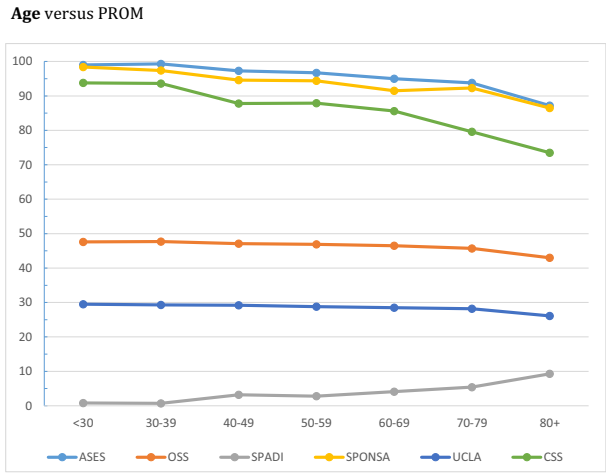

Figure 1 Patient reported-outcome measure (*PROM*) scores compared against age for each PROM investigated. *ASES*, American Shoulder and Elbow Surgeons; *OSS*, Oxford Shoulder Score; *SPADI*, Shoulder Pain and Disability Index; *SPONSA*, Stanmore Percentage of Normal Shoulder Assessment; *UCLA*, University of California, Los Angeles; *CSS*, Constant-Murley Shoulder Score.

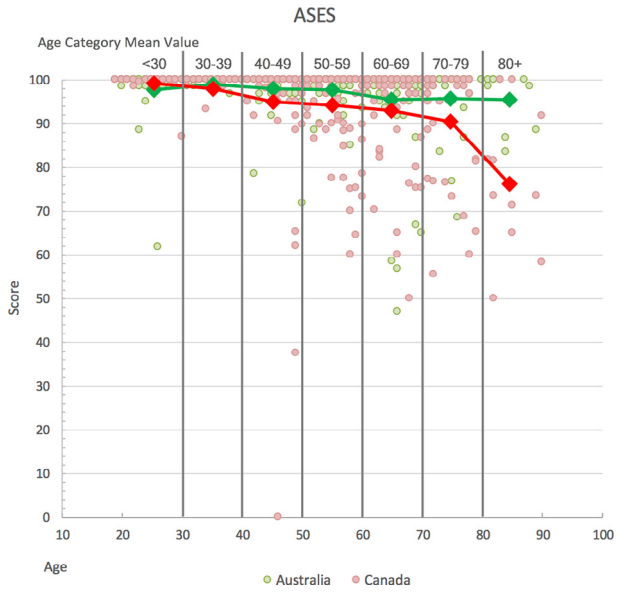

Figure 2 Scattergram and average American Shoulder and Elbow Surgeons (*ASES*) shoulder score by nationality and age group. The data lines are trend lines for the average score per age group. The *red line* represents Canada and the *green line* represents Australia.

(Tables S3, S6, and S7, respectively). Correlation between PROMs is presented in Table VI.

Discussion

It is an important goal to differentiate normal, age-related changes in shoulder function from those changes associated with shoulder pathology; surgical intervention; or implant wear, fatigue, or failure. An important step toward this goal is es-

Figure 3 Scattergram and average Constant-Murley Shoulder Score (*CSS*) by nationality and age group. The data lines are trend lines for the average score per age group. The *red line* represents Canada and the *green line* represents Australia.

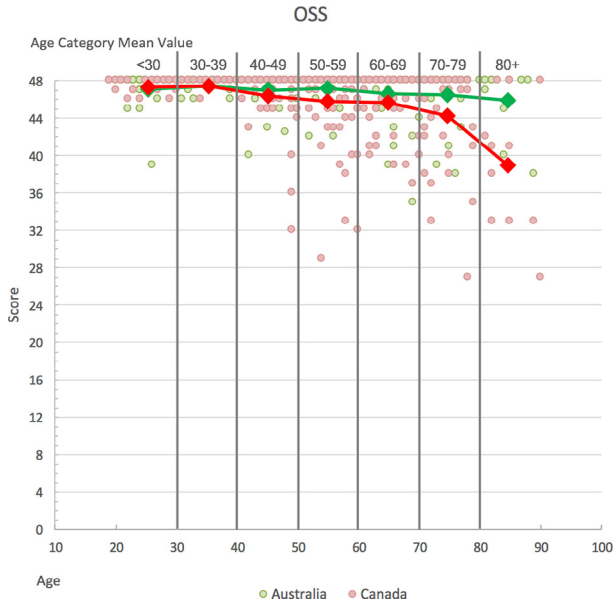

Figure 4 Scattergram and average Oxford Shoulder Score (*OSS*) by nationality and age group. The data lines are trend lines for the average score per age group. The *red line* represents Canada and the *green line* represents Australia.

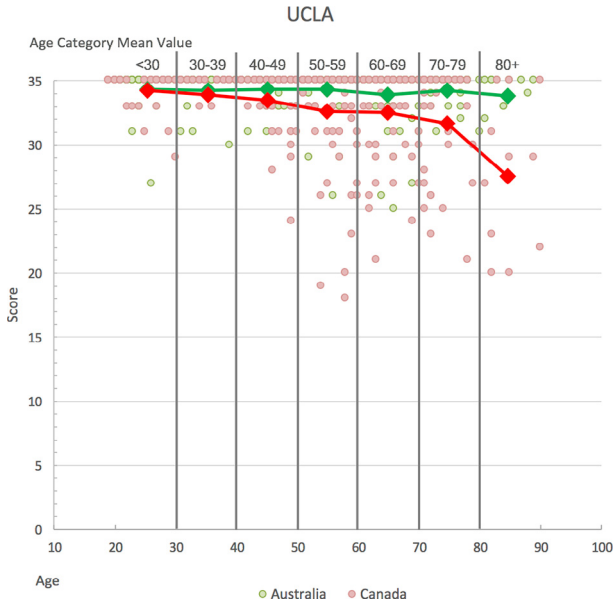

Figure 5 Scattergram and average University of California, Los Angeles (*UCLA*) shoulder score by nationality and age group. The data lines are trend lines for the average score per age group. The *red line* represents Canada and the *green line* represents Australia.

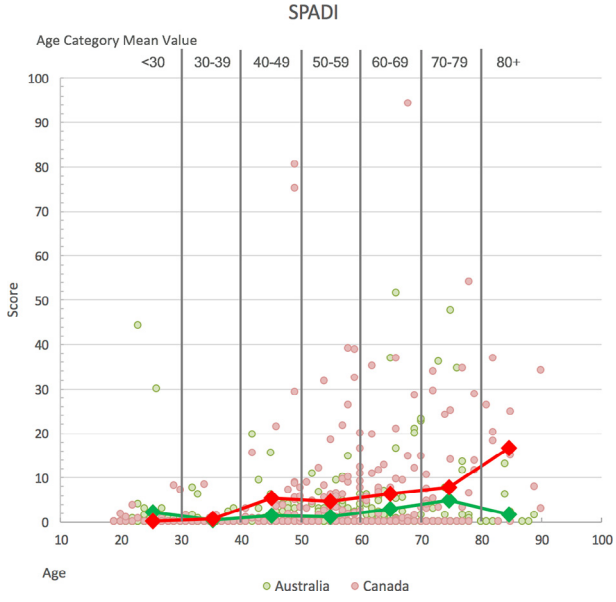

Figure 6 Scattergram and average Shoulder Pain and Disability Index (*SPADI*) by nationality and age group. The data lines are trend lines for the average score per age group. The *red line* represents Canada and the *green line* represents Australia.

tablishing a PROM reference database for individuals without shoulder disease so that clinicians can effectively evaluate the efficacy of treatments in patients on a longitudinal basis.

A recent review identified 44 outcome scores for the shoulder.[21] Many are used inappropriately or modified and not tested for validity, reproducibility, or sensitivity to change. Full validation of a scoring system is essential before it can be recommended for clinical or research use. There remain methodologic inconsistencies and difficulties with some widely used scores.[21]

This study assessed 6 commonly used shoulder PROMs in pathology-free individuals. Of these, the ASES shoulder score, OSS, SPADI, and SPONSA contain subjective-only patient reports of shoulder function. These PROMs have the theoretical advantage of potentially being remotely administered, as they do not require a face-to-face interaction with

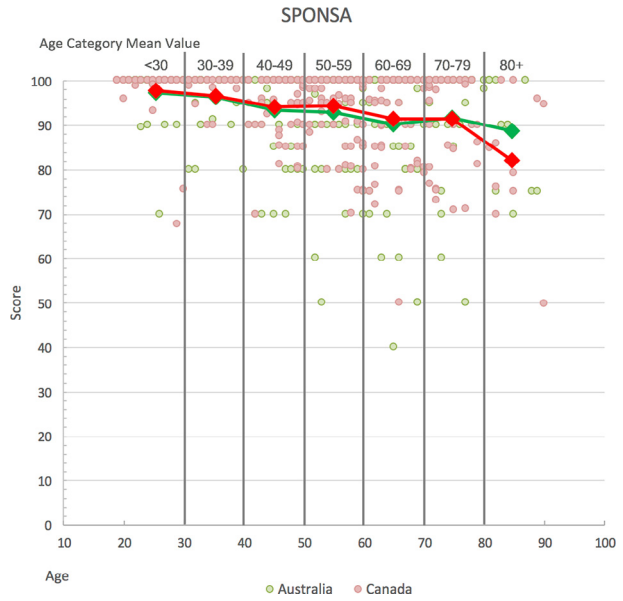

Figure 7 Scattergram and average Stanmore Percentage of Normal Shoulder Assessment (*SPONSA*) by nationality and age group. The data lines are trend lines for the average score per age group. The *red line* represents Canada and the *green line* represents Australia.

an assessor. In comparison, the CSS and UCLA shoulder score have both subjective and clinician-derived objective components to their scoring. These 2 assessment tools require additional time, resources, and personnel to administer (including a face-to-face interaction with an assessor). This study assessed for correlation between the 6 shoulder PROMs, as well as for differences between the PROMs, in relation to sex, age, ethnicity, and geographical location. An important aspect of the analysis was to determine whether there were differences identified between the subjective-only PROMs and the subjective-objective PROMs. Potential benefits exist if there is no reported difference in subjective-only PROMs, as these may be administered remotely, without the need for a face-to-face interaction.

This study reports a statistically significant correlation between all assessment tools, with no difference identified when comparing subjective-only PROMs with combined subjective-objective PROMs. The highest reported correlation was between the OSS (subjective only) and UCLA score (combined subjective and objective). This finding suggests there may be no advantage in using a combined subjective-objective assessment tool. As resource management and cost justification are becoming more of a focus for our institutions, consideration should be given to a shoulder PROM tool that is equivalent but does not require an objective assessment component. Further research is required to determine whether this finding applies to a pathologic shoulder cohort, especially because the collection of objective data requires additional resources, personnel, and expertise to complete.

An association between a history of inactive shoulder pathology and poorer PROM values was seen, with the total numbers of participants having reported such a history being relatively high (15.3%). With increasing age category, the proportion of participants reporting an inactive history increased, as did the proportion of participants with a history of dominant shoulder surgery (>3 years ago) who were asymptomatic (Table VII). Overall, even proportions of Canadian and Australian participants reported inactive shoulder pathology; however, as age increased, more Canadians had a positive history. Although it is logical that with increasing age, more persons would have a history of asymptomatic shoulder pathology, it is possible that a selection bias may have contributed to this finding. Certainly, a history of inactive shoulder pathology should be taken into consideration when choosing an asymptomatic "control" group to compare against an interventional group.

When we compared all PROMs, there was no statistically significant association between PROM score and a current elbow or wrist problem. This finding supports the conclusion that these 6 shoulder PROMs are specific to the shoulder and can discriminate pathology with the shoulder as its primary source from other sources of upper-limb incapacity. Although it is possible that a participant with a current elbow or wrist problem may have had overlapping upper-limb issues that could have influenced some (or all) of his or her responses, the results did not support this. Further research is required to determine whether this finding applies to a pathologic shoulder cohort.

In this study, shoulder data were collected electronically from 2 asymptomatic, distinct, remote, Westernized populations of different geographical locations that were representative of their local populations.[1,11] To our knowledge, this has not been investigated previously. The slightly higher proportion of persons of European descent (white participants by default) represented in the Australian cohort is consistent with that reported by the Australian Bureau of Statistics.[1] The higher proportions of Chinese, Middle Eastern, and Asian Indians represented in the Canadian cohort are consistent with those reported by Statistics Canada.[11] Although there was a difference observed between the cohorts regarding ethnicity, the numbers were too small to allow for any statistical assessment.

A trend for female participants to report similar or poorer overall PROM scores was found for all PROMs assessed. However, not all of the PROMs investigated were statistically significant. This finding is similar to findings of other studies that have reported poorer CSS[6,16] and OSS[6] asymptomatic scores associated with female sex. This finding differs from that of a study assessing asymptomatic shoulder ASES scores in young, active adults, which reported no difference when assessing sex.[3] Our study had a larger distribution of ages and a higher proportion of elderly participants, and this may explain the differences reported in the literature. Further research is required to assess whether this study's finding is a trend or represents a true difference.

Our study found an inverse correlation between age and reported PROM (ie, as age increases, the reported PROM score is poorer). This finding is not surprising given the

Table VI Comparison of patient-reported outcome assessments using Spearman correlation coefficients (*P* value) to assess for correlation

	ASES score	CSS	OSS	UCLA score	SPONSA
SPADI score	−0.78 (<.0001)	−0.41 (<.0001)	−0.78 (<.0001)	−0.68 (<.0001)	−0.63 (<.0001)
SPONSA	0.61 (<.0001)	0.35 (<.0001)	0.53 (<.0001)	0.47 (<.0001)	
UCLA score	0.65 (<.0001)	0.42 (<.0001)	0.80 (<.0001)		
OSS	0.71 (<.0001)	0.44 (<.0001)			
CSS	0.48 (<.0001)				

ASES, American Shoulder and Elbow Surgeons; *CSS*, Constant-Murley Shoulder Score; *OSS*, Oxford Shoulder Score; *UCLA*, University of California, Los Angeles; *SPONSA*, Stanmore Percentage of Normal Shoulder Assessment; *SPADI*, Shoulder Pain and Disability Index.
A correlation coefficient of 0.7 or greater is considered very strong; 0.4-0.69, strong; 0.3-0.39, moderate; 0.2-0.29, weak; and 0.01-0.19, negligible.

Table VII Demographic data: history of shoulder surgery and inactive shoulder problem by age category

	History of shoulder surgery, %		History of inactive shoulder problem, %	
	Australia	Canada	Australia	Canada
<30 y	0.0	0.0	15.4	3.0
30-39 y	0.0	0.0	6.3	3.1
40-49 y	0.0	2.0	13.5	21.6
50-59 y	1.4	2.7	13.7	20.5
60-69 y	1.4	4.3	20.8	20.3
70-79 y	0.0	7.0	9.5	25.6
≥80 y	0.0	18.2	23.1	54.5

age-related changes that occur over time, as well as the accumulated medical and surgical comorbidities that can affect shoulder function. Very few studies exist that have reported associations between shoulder PROMs and age, with a few studies reporting poorer CSS in older age groups.[6,16]

In addition to the inverse correlation between age and reported PROM, a large variation in PROM values was reported in individuals aged 80 years or older. These poorer PROM reports were more pronounced in the Canadian cohort aged 80 years or older compared with their Australian equivalents. This trend is well demonstrated when comparing each PROM graph, where national PROMs become more divergent in individuals aged 80 years or older (Figs. 2-7). This finding may represent a true difference between the Australian and Canadian populations or may represent a selection bias. Certainly, this finding suggests that there are likely many other determinants of health and function that may influence the subjective and objective scores reported and possibly the accuracy of "asymptomatic" values in this older age group.

No difference in OSS values for participants younger than 79 years was found (Table S5). The greatest variation in OSS values was recorded in the group aged 80 years or older, with some respondents in this subgroup recording OSS values of 48 out of 48 (even up to 90 years [OSS range, 27-48 out of 48]). Overall, this study suggests that PROM values are highly variable in individuals aged 80 years or older, making comparisons in this older age group less reliable. Care should be taken when interpreting these data and applying generalizations to different populations in geographically distinct locations (especially in individuals aged ≥80 years).

There were comparable CSS and SPONSA values between the national cohorts. This finding suggests that future studies using the CSS and/or SPONSA and a control group can be more easily generalized against populations, independent of the study country of origin. Further studies need to be completed to determine whether this principle applies to other countries that use an electronic database.

Our study found a statistically significant difference between the Canadian and Australian cohorts for the ASES, UCLA, and SPADI scores. This finding is difficult to interpret and may reflect a true difference between the population cohorts—or a selection or reporting bias. We recommend that future studies using the ASES, UCLA, and SPADI scores should perform comparisons against a control group sourced from the same country of origin and be age and sex matched.

To our knowledge, this study represents the largest database of nonpathologic shoulder PROM values reported in the literature. Other researchers have reported asymptomatic values for some of the PROMs examined in this study, but few have approached the numbers collected in this study or the participants were limited to young, active individuals.[3,4,6,16,24,26] To our knowledge, there exist no data in the literature reporting nonpathologic shoulder PROM values for the SPADI or SPONSA shoulder scores.

Several studies have reported that the theoretically best shoulder score may not be reported in an asymptomatic and disease-free population.[2,3,14,16,24,26] However, very few have examined the effect of age, sex, hand dominance, race, or geographical location on shoulder PROM scores in pathology-free individuals. No other study has compared 6 shoulder PROMs in "asymptomatic" individuals of all ages.

This 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 randomization. The primary benefit of randomization is 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 randomization method of participant identification was used. Another potential source of selection bias involves the use of electronic questionnaires, in which case participants may have declined to be involved because of 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 inactive shoulder pathology may have chosen not to participate in the study, citing that their shoulder was not "normal." Although we chose to exclude participants with active shoulder disease, we did include participants with a history of a previous shoulder 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 shoulder problem may have had 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 shoulder pathology. As no radiographs or ultrasound scans were taken to confirm whether participants had asymptomatic shoulder disease, it is possible that some of these individuals were included in the cohorts.

The adoption of electronic, automated PROM data collection relies on the assumption that a patient prefers this virtual method of follow-up in preference to a face-to-face clinician-patient interaction. It also assumes that a patient is capable of and willing to be actively participating independently in his or her management. It also requires the patient to be computer and E-mail savvy, an assumption that may not be correct for the entire population. This method of electronic data collection has the potential advantage of allowing remote follow-up and, for research purposes, enables data collection at an international level.

Conclusion

An electronic PROM data collection system can be used effectively across different continents. Differences in sex, age, and geographical location will affect PROM shoulder scores in pathology-free individuals and should be taken into consideration when PROMs are being used to compare patient outcomes. This study has established normative values for the ASES shoulder score, CSS, OSS, UCLA shoulder score, SPADI score, and SPONSA. Future studies assessing a pathologic patient cohort should perform comparisons against a sex- and age-matched control cohort, ideally sourced from their same country of origin.

Disclaimer

The individual research tool registrations were paid for by OBERD Pty Ltd and are included as part of the licensing agreement.

The authors, their immediate families, and any research foundations with which they are affiliated have not received any financial payments or other benefits from any commercial entity related to the subject of this article.

Acknowledgments

We acknowledge Mami Okada, University of British Columbia, for her help in the establishment of the study in Canada; Tara-Louise McLean, research assistant to J.M.M., for her help in the establishment of the study in Australia and Canada; Suzanne Edwards, statistician, Data Management and Analysis Centre at the University of Adelaide, for her help with the statistical analysis of the collected data; Tiffany Gill, University of Adelaide, for her help with the study design; and Catherine Hill, University of Adelaide, for her help with the study design.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jse.2017.08.016.

References

1. Australian Bureau of Statistics. Cultural diversity in Australia. Vol ABS catalogue number 2071.0. Canberra. <http://www.abs.gov.au/ausstats/abs@.nsf/mf/2071.0>; 2011, accessed September 30, 2016.
2. Brinker MR, Cuomo JS, Popham GJ, O'Connor DP, Barrack RL. An examination of bias in shoulder scoring instruments among healthy collegiate and recreational athletes. J Shoulder Elbow Surg 2002;11:463-9. http://dx.doi.org/10.1067/mse.2002.126209
3. Clarke MG, Dewing CB, Schroder DT, Solomon DJ, Provencher MT. Normal shoulder outcome score values in the young, active adult. J Shoulder Elbow Surg 2009;18:424-8. http://dx.doi.org/10.1016/j.jse.2008.10.009
4. Clement ND, Court-Brown CM. Oxford shoulder score in a normal population. Int J Shoulder Surg 2014;8:10-4. http://dx.doi.org/10.4103/0973-6042.131849
5. Conboy VB, Morris RW, Kiss J, Carr AJ. An evaluation of the Constant-Murley shoulder assessment. J Bone Joint Surg Br 1996;78:229-32.
6. Constant CR, Gerber C, Emery RJ, Søjbjerg JO, Gohlke F, Boileau P. A review of the Constant score: modifications and guidelines for its use. J Shoulder Elbow Surg 2008;17:355-61. http://dx.doi.org/10.1016/j.jse.2007.06.022
7. Coons SJ, Gwaltney CJ, Hays RD, Lundy JJ, Sloan JA, Revicki DA, et al. Recommendations on evidence needed to support measurement equivalence between electronic and paper-based patient-reported outcome (PRO) measures: ISPOR ePRO Good Research Practices Task Force report. Value Health 2009;12:419-29. http://dx.doi.org/10.1111/j.1524-4733.2008.00470.x

8. Dawson J, Fitzpatrick R, Carr A. The assessment of shoulder instability. The development and validation of a questionnaire. J Bone Joint Surg Br 1999;81:420-6.
9. Dawson J, Fitzpatrick R, Carr A. Questionnaire on the perceptions of patients about shoulder surgery. J Bone Joint Surg Br 1996;78:593-600.
10. Dawson J, Rogers K, Fitzpatrick R, Carr A. The Oxford shoulder score revisited. Arch Orthop Trauma Surg 2009;129:119-23. http://dx.doi.org/10.1007/s00402-007-0549-7
11. Government of Canada. Ethnic origins for Canada, provinces and territories Calgary. <http://www12.statcan.ca/census-recensement/2011/as-sa/98-310-x/98-310-x2011001-eng.cfm>; 2011, accessed September 30, 2016.
12. Greenhalgh J. The applications of PROs in clinical practice: what are they, do they work, and why? Qual Life Res 2009;18:115-23. http://dx.doi.org/10.1007/s11136-008-9430-6
13. Griffiths-Jones W, Norton MR, Fern ED, Williams DH. The equivalence of remote electronic and paper patient reported outcome (PRO) collection. J Arthroplasty 2014;29:2136-9. http://dx.doi.org/10.1016/j.arth.2014.07.003
14. Hunsaker FG, Cioffi DA, Amadio PC, Wright JG, Caughlin B. The American academy of orthopaedic surgeons outcomes instruments: normative values from the general population. J Bone Joint Surg Am 2002;84:208-15.
15. James MA. Use of the Medical Research Council muscle strength grading system in the upper extremity. J Hand Surg Am 2007;32:154-6. http://dx.doi.org/10.1016/j.jhsa.2006.11.008
16. Katolik LI, Romeo AA, Cole BJ, Verma NN, Hayden JK, Bach BR. Normalization of the Constant score. J Shoulder Elbow Surg 2005;14:279-85. http://dx.doi.org/10.1016/j.jse.2004.10.009
17. Kocher MS, Horan MP, Briggs KK, Richardson TR, O'Holleran J, Hawkins RJ. Reliability, validity, and responsiveness of the American Shoulder and Elbow Surgeons subjective shoulder scale in patients with shoulder instability, rotator cuff disease, and glenohumeral arthritis. J Bone Joint Surg Am 2005;87:2006-11. http://dx.doi.org/10.2106/JBJS.C.01624
18. MacDermid JC, Solomon P, Prkachin K. The Shoulder Pain and Disability Index demonstrates factor, construct and longitudinal validity. BMC Musculoskelet Disord 2006;7:12. http://dx.doi.org/10.1186/1471-2474-7-12
19. Maréchal E. Ruptures degeneratives de la coiffe des rotateurs de l'épaule. Evaluation fonctionnelle resultats du traitement chirurgical. Lyon, France: Faculté de Médecine Grange, Université Claude Bernard; 1990.
20. Mitchell K, Gutierrez SB, Sutton S, Morton S, Morgenthaler A. Reliability and validity of goniometric iPhone applications for the assessment of active shoulder external rotation. Physiother Theory Pract 2014;30:521-5. http://dx.doi.org/10.3109/09593985.2014.900593
21. Noorani AM, Roberts DJ, Malone AA, Waters TS, Jaggi A, Lambert SM, et al. Validation of the Stanmore Percentage of Normal Shoulder Assessment. Int J Shoulder Surg 2012;6:9-14. http://dx.doi.org/10.4103/0973-6042.94307
22. Richards RR, An KN, Bigliani LU, Friedman RJ, Gartsman GM, Gristina AG, et al. A standardized method for the assessment of shoulder function. J Shoulder Elbow Surg 1994;3:347-52.
23. Roach KE, Budiman-Mak E, Songsiridej N, Lertratanakul Y. Development of a shoulder pain and disability index. Arthritis Care Res 1991;4:143-9.
24. Sallay PI, Reed L. The measurement of normative American Shoulder and Elbow Surgeons scores. J Shoulder Elbow Surg 2003;12:622-7. http://dx.doi.org/10.1016/S1058-2746(03)00209-X
25. Smith MJ, Marberry KM. Reliability of a novel, web-based, shoulder-specific, patient-reported outcome instrument. Curr Orthop Pract 2013;24:64-7. http://dx.doi.org/10.1097/BCO.0b013e3182781c05
26. Soldatis JJ, Moseley JB, Etminan M. Shoulder symptoms in healthy athletes: a comparison of outcome scoring systems. J Shoulder Elbow Surg 1997;6:265-71.

CHAPTER 5 – ASYMPTOMATIC REFERENCE VALUES FOR THE DASH AND PRWHE – ELECTRONIC DATA COLLECTION AND ITS CLINICAL IMPLICATIONS

As published in Journal of Hand Surgery: European Volume. 2018. Nov; 43 (9): 988-993.

Introduction

The efficacy of medical and surgical treatments for wrist and hand disorders, are commonly assessed using Patient Reported Outcome Measures (PROMs). Several wrist and hand PROMs have been described and validated, with the Disability of Arm, Shoulder and Hand (DASH) questionnaire [12] and Patient-Rated Wrist/Hand Evaluation (PRWHE) [11], being investigated in this thesis. Accurate interpretation of PROMs requires an understanding of patient factors that have the potential to impact scores; such as previous or current injury or surgery, age, gender, handedness, ethnicity and geographical location.

Aims

The aim of this study was to assess whether the DASH and PRWHE scores were comparable in normal, healthy, pathology-free individuals of different age, gender, ethnicity, handedness, history of pathology and nationality. The purpose of this study was to establish a data bank of "normal" population values for the DASH and PRWHE upper limb scores. The hypothesis was that there is no difference in reported DASH and PRWHE values in an asymptomatic "normal" population, when comparing sub-groups of differing age, gender, ethnicity, handedness, history of pathology and different nationality.

Main Findings

This study of 584 participants represents the only available data on normative PRWHE values in the literature. It also represents one of the largest databases of normative DASH values reported in the literature [73-75]. No other study has compared upper limb PROM scores from normal, pathology-free individuals from international cohorts, or collected them remotely and electronically via the same research database.

This study found that there was an association between the DASH and age and the PRWHE and age, demonstrating that as age increased, reported PROM scores were poorer.

Females reported poorer DASH and PRWHE scores compared to males, which was more evident as age increased. This finding was consistent with previous studies that reported poorer DASH scores in asymptomatic, "normal" females [74, 76, 77]. Poorer PRWHE scores for females has not previously been reported in an asymptomatic cohort.

There was no difference in the two reported PROM scores when considering handedness or geographical location. Twelve percent of participants reported *a history of an inactive wrist or hand problem*, which was associated with 10.5% poorer reported PROM score, compared to those who had no such history. This seems logical and suggests that *a history of an inactive wrist or hand problem* should be taken into consideration when choosing an asymptomatic "control" group to compare against an interventional group.

Participants who reported *a history of an active shoulder or elbow problem*, reported poorer PROM scores compared to those who had no such history. This finding supports the contention that there may be considerable overlap in the functional questions contained in the two PROMs investigated. This suggests that the DASH and PRWHE cannot discriminate a primarily wrist- or hand-source of pathology from other sources of upper limb incapacity. This should be considered when using these PROMs in patients with multiple, concurrent upper limb pathologies

Future directions

Studies using an electronic control group should consider differences in gender, age, ethnicity, handedness, history of pathology and nationality, when using wrist and hand PROMs to assess patient outcomes. This study has established an electronic, "normal" control group for studies using the DASH and PRWHE PROMs. When using these PROMs, the control group can be sourced from a pre-established control group within an existing database, without necessarily being sourced from the same country of origin. Care should be taken when using the DASH and PRWHE PROMs in patients with multiple, concurrent upper limb pathologies, or in patients with *a history of an inactive wrist or hand problem*, as these factors are likely to impact on the PROM values reported.

Statement of Authorship & Contribution

This section provides a statement of the contribution of each author for all the peer review publications in the thesis. All co-authors have signed the statement of contribution.

Signed statements follow.

Statement of Authorship

Title of Paper	Asymptomatic Reference Values for the DASH and PRWHE - Electronic Data Collection and Its Clinical Implications
Publication Status	☐ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted w ork w ritten in manuscript style
Publication Details	Journal of Hand Surgery: European Volume. 2018. In-press. JHSE-D-16-00541R5

Principal Author

Name of Principal Author (Candidate)	Dr James Marcus McLean		
Contribution to the Paper	Conceptual project design Etthics committee applications (Canada & Australia) Data collection (oversight & manual collection) Data collation & processing Manuscript preparation & submission Editor responses & preparing of re-submissions		
Overall percentage (%)	80%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	23 May 2018

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Dr Afsana Hasan		
Contribution to the Paper	Data collation, processing & analyses Presentation of findings Manuscript preparation & Editor responses		
Signature		Date	03 June 2018

Name of Co-Author	Dr Jake Willet		
Contribution to the Paper	Data collection (Australia) Pre-liminary data collation, processing & analyses Presentation of preliminary data		
Signature		Date	30 May 2018

Name of Co-Author	Dr Matthew Jennings		
Contribution to the Paper	Data collection (Australia) Pre-liminary data collation, processing & analyses Presentation of preliminary data		
Signature		Date	04 June 2018

Name of Co-Author	Dr Kimberly Brown		
Contribution to the Paper	Data collation, processing & analyses Manuscript preparation		
Signature		Date	02 June 2018

Name of Co-Author	Dr Laura Goodwins		
Contribution to the Paper	Data collation, processing & analyses Manuscript preparation		
Signature		Date	30 May 2018

Name of Co-Author	Dr Tom Goetz		
Contribution to the Paper	Project oversight (Canada) Etthics committee application (Canada)		
Signature		Date	30 May 2018

Published Manuscript

ASYMPTOMATIC REFERENCE VALUES FOR THE DASH AND PRWHE – ELECTRONIC DATA COLLECTION AND ITS CLINICAL IMPLICATIONS

Journal of Hand Surgery: European Volume. 2018. Nov; 43 (9): 988-993.

Full Length Article

JHS(E)

Journal of Hand Surgery
(European Volume)
2018, Vol. 43(9) 988–993

DOI: 10.1177/1753193418793027
journals.sagepub.com/home/jhs
SAGE

Asymptomatic reference values for the Disability of Arm, Shoulder and Hand and Patient-Rated Wrist/Hand Evaluation – electronic data collection and its clinical implications

James M. McLean[1,2], Afsana P. Hasan[2], Jake Willet[2], Matthew Jennings[2], Kimberly Brown[3], Laura Goodwins[3] and Tom Goetz[1]

Abstract

The purpose of this study was to establish normal asymptomatic population values for the Disability of Arm, Shoulder and Hand and Patient-Rated Wrist/Hand Evaluation in healthy, asymptomatic individuals of different age, gender, ethnicity, handedness and nationality, using electronic data collection. Two-hundred and ninety-two Australian and 293 Canadian citizens with no active wrist pain, injury or pathology in their dominant hand, were evaluated. Participants completed an electronically administered questionnaire and were assessed clinically. There was no statistically significant association between both wrist scores and nationality. There was a statistically significant association between both wrist scores and age, demonstrating that as age increased, normal wrist function declined. This study has established an electronic, asymptomatic control group for future studies using these scores. When using the Disability of Arm, Shoulder and Hand and Patient-Rated Wrist/Hand Evaluation, the control group can be sourced from a pre-established control group within a database, without necessarily being sourced from the same country of origin.

Level of evidence: II

Keywords

Disability of Arm, Shoulder and Hand, Patient-Rated Wrist/Hand Evaluation, wrist, hand, normal, reference

Date received: 17th November 2016; revised: 7th July 2018; accepted: 15th July 2018

Introduction

The efficacy of current and future medical and surgical treatments for wrist and hand disorders can be evaluated with outcome assessments tools. Pre- and post-operative Patient-Reported Outcome Measures (PROMs) are important tools in assessing a patient's suitability for surgery, expected outcome and post-operative recovery. PROMs data can be collected and processed electronically, enabling faster processing with minimal clinician input (Griffiths-Jones et al., 2014). Accurate interpretation of PROMs requires an understanding of patient factors that have the potential to impact scores, such as previous or current injury, age, gender, handedness, ethnicity and nationality.

The Disability of Arm, Shoulder and Hand (DASH) questionnaire and Patient-Rated Wrist/Hand Evaluation (PRWHE) were chosen for this study because they both demonstrate good reliability, validity and responsiveness (Beaton et al., 2001;

[1]Department of Orthopaedics, University of British Columbia, Vancouver, British Columbia, Canada
[2]Discipline of Orthopaedics and Trauma, University of Adelaide, North Terrace, Adelaide, Australia
[3]Discipline of Medicine, University of Adelaide, North Terrace, Adelaide, Australia

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

Changulani et al., 2008; Gay et al., 2003; Hoang-Kim et al., 2011; MacDermid, 2001; MacDermid et al., 2000).

The purpose of this study was to establish asymptomatic population values for the DASH and PRWHE, using an electronic data collection system. These novel data systems will support more accurate interpretations of outcome assessments and allow for direct comparison to age- and gender-matched individuals, or a comparison with individuals with specific characteristics.

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

Methods

Patient selection

With prior Research Ethics Board approval, between November 2014 to October 2015, healthy volunteers were recruited in Australia and Canada from a variety of sources, including Drivers Licensing Offices, Medical Outpatients Facilities and various community centres. There were no study advertisements or incentives and participants were not paid.

The inclusion criteria included: English-speaking adults with no reported active wrist or hand pathology in either hand. Participants were directed to answer their questionnaires in relation to their dominant hand. Potential participants self-reported a history of dominant hand pain or pathology; no medical charts or radiographs were reviewed to categorize asymptomatic participants.

Exclusion criteria included: cognitive impairment; a history of inflammatory or wrist arthritis; significant cervical spine problems that interfered with their function; active wrist or hand pathology, surgical arthrodesis or arthroplasty; or wrist or hand surgery within the past 3 years.

A history of inactive wrist or hand pathology, including previous surgery, was recorded. A history of active shoulder and elbow pathology was recorded.

Data collection

Participants self-administered 45 questions (DASH 30 questions and PRWHE 15 questions), using a web-based data collection tool (OBERD, Universal Research Solutions, Columbia, MO, USA), on an electronic mobile device (tablet or laptop computer). Data handling by the recruiters was minimal, ensuring that the investigators were partially blinded to the participants' results.

Primary outcome measures

The DASH score is a 30-item upper extremity outcome assessment tool that assesses an individual's symptoms and functional status (Hudak et al., 1996). Each item is scored between 1–5, and the maximum score of 100 correlates with a high disability.

The PRWHE score is a 15-item wrist and hand assessment tool that assesses an individual's wrist and hand pain (5 items) and function (10 items) (MacDermid et al., 1998). Each item is scored between 0–10 and the functional score is divided by 2. The maximum score of 100 correlates with a high disability.

Statistical analysis

A power calculation was performed to determine the sample size necessary to detect a clinically significant difference in wrist scores of 20% at a power of 80% and an alpha value of 0.05 ($n = 280$). Associations between nationality, 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 wrist 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. Wrist scores were therefore considered to be counts. Poisson regressions were performed and ranged from 0.0124 to 3.1994. Confidence interval (CI) was set at 95% for a two-way mixed effects model and absolute agreement. Initially country cohort and all confounders were included in a multivariable Poisson regression model for each wrist score outcome variable. Backwards stepwise elimination was then performed until all covariates had a p value <0.2.

Results

Two-hundred and ninety-two Australian and 293 Canadian citizens with no active wrist pain, injury or pathology in their dominant hand completed the questionnaire. Demographics of the cohorts are presented in Tables 1 and 2. The effect of a history of inactive wrist/hand problem or an active shoulder and/or elbow problem on a participant's PROM scores are presented in Table 3.

DASH

There was a statistically significant association between the DASH and age ($p < 0.0001$, EE = 0.06, 95% CI = 0.02, 0.22). For every 1-year increase in

Table 1. A comparison of ethnicity of the two international cohorts.

Ethnicity	Australia ($n = 291$)	Canada ($n = 293$)	Total ($n = 584$)
Caucasian	258 (88.7%)	217 (74.0%)	276 (83.4%)
Chinese	9 (3.1%)	20 (6.8%)	29 (4.9%)
Asian Indian	6 (2.1%)	18 (6.1%)	24 (4.1%)
Middle Eastern	5 (1.7%)	17 (5.8%)	22 (3.7%)
Other Asian	5 (1.7%)	9 (3.1%)	14 (2.4%)
Filipino	5 (1.7%)	8 (2.7%)	13 (2.2%)
Indigenous	2 (<1%)	3 (1.1%)	5 (<1%)
African American	1 (<1%)	1 (<1%)	2 (<1%)

Table 2. A comparison of demographics of the two international cohorts.

	Australian cohort ($n = 292$)	Canadian cohort ($n = 293$)	Total ($n = 585$)	Comparing Australian and Canadian cohorts
Male	140 (48.9%)	140 (47.8%)	280 (47.8%)	$p = 0.10$
Female	152 (52.1%)	153 (52.2%)	305 (52.2%)	
Left	29 (9.9%)	30 (10.2%)	59 (10.1%)	$p = 0.86$
Right	263 (90.1%)	263 (89.8%)	526 (89.9%)	
Age < 30	35 (12%)	33 (11.2%)	68 (11.6%)	$p = 0.84$
Age 30–39	34 (11.9%)	37 (12.6%)	71 (12.1%)	
Age 40–49	50 (17.1%)	51 (17.4%)	101 (17.3%)	
Age 50–59	62 (21.2%)	65 (22.2%)	127 (21.7%)	
Age 60–69	66 (22.6%)	65 (22.2%)	131 (22.4%)	
Age 70–79	23 (7.8%)	22 (7.5%)	45 (7.7%)	
Age 80+	18 (6.1%)	20 (6.8%)	38 (6.5%)	
Average age	53 (range 18–94)	53 (range 18–96)	53 (range 18–96)	
Privately insured	148 (50.7%)	0 (0.0%)		
Publicly insured	144 (49.3%)	293 (100.0%)		
Participant reported a history of an inactive (previous) wrist or hand problem	36 (12.1%)	38 (12.9%)	74 (12.6%)	$p = 0.77$
Participant reported a history of an active shoulder or elbow problem	18 (6.1%)	29 (9.8%)	47 (8%)	$p = 0.09$

age, the mean DASH value increased by 3.3% ($p < 0.0001$; Table 4; Supplemental Figure S1). There was no statistically significant association between the DASH and handedness ($p = 0.5841$) or nationality (adjusting for age ($p = 0.8790$) and adjusting for gender ($p = 0.7715$)). Australians reported DASH values 1.3% lower than Canadians, which was not statistically significant. There was a statistically significant association between DASH and gender ($p < 0.0001$, EE = 2.41, 95% CI = 1.65, 3.52). The mean DASH value for females was 7.8 compared with 3.2 for males, with females reporting a mean DASH score 2.4 times that of males.

PRWHE

There was a statistically significant association between the PRWHE and age ($p < 0.0001$, EE = 0.53, 95% CI = 0.38, 0.72). For every 1-year increase in age, the mean PRWHE value increased by 2.3% ($p < 0.0001$; Table 5; Supplemental Figure S2). There was no statistically significant association

Table 3. Average DASH and PRWHE scores in participants with/without a history of an inactive wrist/hand problem and participants with/without an active shoulder/elbow problem.

	DASH	PRWHE
History of an inactive wrist/ hand problem	10.0	11.4
No history of a wrist/hand problem	4.9	5.6
Mean difference	5.1[a]	5.8[b]
Active shoulder and/or elbow problem	21.3	21.1
No active shoulder/elbow problem	4.2	5.1
Mean difference	17.1[c]	16[d]

[a]EE = 2.05, 95% CI = 1.55, 2.71, $p < 0.01$.
[b]EE = 2.05, 95% CI = 1.52, 2.75, $p < 0.01$.
[c]EE = 5.01, 95% CI = 4.04, 6.22, $p < 0.001$.
[d]EE = 4.13, 95% CI = 3.25, 5.27, $p < 0.001$.
(DASH) Disability of Arm, Shoulder and Hand; (PRWHE) Patient-Rated Wrist/Hand Evaluation.

Table 4. Mean DASH scores by age for Australian and Canadian cohorts, and the combined cohorts.

Age	Australia	Canada	Combined
<30	1.44	0.49	0.97
30–39	2.00	1.09	1.54
40–49	3.31	2.02	2.67
50–59	3.94	5.23	4.60
60–69	4.88	6.37	5.63
70–79	10.60	15.99	13.30
80+	20.64	27.68	24.35
Overall	4.99	6.22	5.61

DASH (lowest (best) score 0; highest (worst) score = 100).

Table 5. Mean PRWHE scores by age for Australian and Canadian cohorts, and the combined cohorts.

Age	Australia	Canada	Combined
<30	6.59	3.87	1.19
30–39	4.68	4.58	2.53
40–49	10.03	5.90	3.90
50–59	7.76	6.36	5.24
60–69	9.78	9.78	7.88
70–79	9.98	17.70	14.80
80+	15.18	17.46	18.80
Overall	5.94	6.79	6.36

PRWHE (lowest (best) score 0; highest (worst) score = 100).

between PRWHE and handedness ($p = 0.4049$) or nationality (adjusting for age ($p = 0.6769$) and adjusting for gender ($p = 0.7369$)). Australians reported PRWHE values 0.4% higher than Canadians, which was not statistically significant. There was a statistically significant association between PRWHE and gender ($p < 0.0001$, EE = 2.1, 95% CI = 1.5, 3.1). The mean PRWHE value for females was 8.6 compared with 4.1 for males, with females reporting a mean PRWHE score 2.1 times that of males.

Discussion

This study established a reference database of normative values for individuals without wrist or hand conditions, to evaluate the efficacy of surgical, non-surgical and hand therapy interventions longitudinally. DASH and PRWHE values were comparable between national cohorts (Tables 4 and 5). This suggests that future international studies using these tools can be performed using a combined control group, without necessarily needing to be sourced from the same country of origin as the proposed study. This can enable larger, combined sample sizes to be used in the assessment of future interventions.

In this study, data was collected electronically from two normal, distinct, remote, Westernized populations of different countries, that were representative of their local populations (Australian Bureau of Statistics, 2011; Canada Statistics, 2011).

The DASH and PRWHE assessment tools were chosen, as they are both subjective scores, allowing them to be administered remotely and electronically at low cost.

A positive relationship was observed between age and the DASH and PRWHE score reported, indicating that as age increases, the degree of baseline, non-pathological disability increases. This finding was consistent with other PROM studies (Aasheim and Finsen, 2014; Hunsaker et al., 2002; Jester et al., 2005; Klum et al., 2012) and likely indicates that age-related changes (i.e. degeneration and loss of strength) and the accumulation of medical and surgical comorbidities influence upper limb function. Among the younger population (aged below 30), the average scores were not as close to perfect as expected. This finding suggests that an excellent DASH or PRWHE score may not reflect a realistic or achievable goal in each individual patient (Clarke et al., 2009) and a considered understanding of other factors that can influence function are required for a more accurate interpretation.

There was a significantly poorer reported score among females compared with males across both PROM tools. This has been previously described only for the DASH in the general population (Aasheim and Finsen, 2014; Hunsaker et al., 2002; Jester et al., 2005).

Table 6. Normative DASH values reported in the literature.

Author, year	*N*	Cohort	Mean
Hunsaker et al., 2002	1706	Randomly selected households in the United States	10
Aasheim and Finsen, 2014	1000	Randomly selected adults aged 20–80 years in Norway	13
Klum et al., 2012	750	Healthy, working volunteers aged 18–65 years in Germany	9
Jester et al., 2005	716	Employed adults in Germany	13

This study reported an association between a history of inactive wrist or hand pathology and PROM score (Table 3). This seems logical, and as such, should be taken into consideration when choosing a control group to compare against an interventional group.

This study reported an association between a history of an active shoulder and/or elbow problem and PROM score. As the DASH assesses the overall function of the upper limb, this finding was expected for the DASH. However, for the PRWHE, there is considerable overlap in the functional questions, which suggests that it may not isolate wrist- and hand-specific function and may fail to discriminate a primarily wrist- and hand-source of pathology from other sources of upper limb incapacity. This should be considered when using the PRWHE in patients with multiple, concurrent upper limb pathologies.

This study represents one of the largest databases of combined asymptomatic population DASH and PRWHE values reported in the literature. Thus far, no other study has compared normal scores in international cohorts or collected them remotely and electronically via the same research database. In the literature, normative values have been collected for the DASH (Table 6); these studies combined both healthy participants with participants with disabilities, likely resulting in a considerably higher average mean score to that which was reported in this study. There is no available data on normative PRWHE values.

Previous studies have administered their questionnaires via mail, resulting in a variable response rate ranging from 50% to 67% (Aasheim and Finsen, 2014; Hunsaker et al., 2002). An electronic database facilitates much easier data collection at an international level. It enables easy and remote access of data to clinical researchers, simultaneous data entry and analysis, the linking of research PROMs to an electronic medical record, the automated setting of participant reminders and the setting of response time points. It is more convenient for patients as timing is flexible and there is no need to return the survey by post. However, electronically administered questionnaires require respondents to possess a basic level of computer skills; an assumption that may not be correct for all members of the public (Snyder et al., 2012). Patients with severe wrist and hand pathology may also be less willing to complete online questionnaires via computers or tablets and may require help to complete the surveys.

Limitations

As with any observational study, there is the potential for selection bias. Although individuals were approached randomly in this study, no specific randomized method of recruitment was employed. A potential source of procedural bias was the use of electronic questionnaires, and elderly participants may have declined to be involved due to the use of technology. Radiographs were not performed to rule out asymptomatic degenerative wrist disease and it is possible that some individuals with asymptomatic pathology were included. It is assumed that a small percentage of the general population will have asymptomatic or incidental undiagnosed wrist/hand pathology. The Authors feel that the carefully chosen inclusion and exclusion criteria in this study produced an asymptomatic population who identified themselves as having 'normal' bilateral wrists and hands.

Participants with a history of dominant hand pathology treated by non-arthroplasty/arthrodesis surgery >3 years ago and no longer experiencing symptoms were included. It is possible that some of these participants still suffered from some slight impairment; however, we only included those who no longer perceived their condition to be a problem.

Acknowledgements Mami Okada, Research Assistant, University of British Columbia, for her help in the establishment of the study in Canada. Tara-Louise McLean, Research Assistant to Dr James McLean, for her help in the establishment of the study in Australia and Canada. Suzanne Edwards, Statistician, Data Management and Analysis Centre at the University of Adelaide, for her help with the statistical analysis of the collected data. Tiffany Gill, University of Adelaide, for her help with the study design. Catherine Hill, University of Adelaide, for her help with the study design.

Declaration of conflicting interests The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding The authors disclosed receipt of the followingfinancial support for the research, authorship, and/or publication of this article: The individual research tool registrations were paid for by OBERD Pty Ltd and are included as part of the licensing agreement.

Ethical approval This study was approved by the Royal Adelaide Hospital Human Research Ethics Committee (ref no. HREC/14/RAH/325) and UBC-Providence Health Care Research Ethics Board (ref no. H14-02030).

Supplemental material Supplemental material is available at: journals.sagepub.com/doi/suppl/10.1177/1753193418793027

References

Aasheim T, Finsen V. The DASH and the QuickDASH instruments. Normative values in the general population in Norway. J Hand Surgery Eur. 2014, 39: 140–4.

Australian Bureau of Statistics. Cultural diversity in Australia. Reflecting a nation: Stories from the 2011 Census. Canberra: 2011.

Beaton D, Katz J, Fossel A, Wright J, Tarasuk V, Bombardier C. Measuring the whole or the parts? Validity, reliability, and responsiveness of the DASH outcome measure in different regions of the upper extremity. J Hand Ther. 2001, 14: 128–46.

Canada Statistics. 2011 National Household Survey – Ethnic origins for Canada, provinces and territories. Calgary: 2011.

Changulani M, Okonkwo U, Keswani T, Kalairajah Y. Outcome evaluation measures for wrist and hand—which one to choose? Int Orthop. 2008, 32: 1–6.

Clarke MG, Dewing CB, Schroder DT, Solomon DJ, Provencher MT. Normal shoulder outcome score values in the young, active adult. J Shoulder Elbow Surg. 2009, 18: 424–8.

Gay RE, Amadio PC, Johnson JC. Comparative responsiveness of the disabilities of the arm, shoulder, and hand, the carpal tunnel questionnaire, and the SF-36 to clinical change after carpal tunnel release. J Hand Surg Am. 2003, 28: 250–4.

Griffiths-Jones W, Norton M, Fern E, Williams D. The equivalence of remote electronic and paper patient reported outcome (PRO) collection. J Arthroplasty. 2014, 29: 2136–9.

Hoang-Kim A, Pegreffi F, Moroni A, Ladd A. Measuring wrist and hand function: common scales and checklists. Injury. 2011, 42: 253–8.

Hudak PL, Amadio PC, Bombardier C. Development of an upper extremity outcome measure: the DASH (Disabilities of the Arm, Shoulder, and Hand). Am J Ind Med. 1996, 29: 602–8.

Hunsaker FG, Cioffi DA, Amadio PC, Wright JG, Caughlin B. The American academy of orthopaedic surgeons outcomes instruments: normative values from the general population. J Bone Joint Surg Am. 2002, 84-A: 208–15.

Jester A, Harth A, Germann G. Measuring levels of upper-extremity disability in employed adults using the DASH Questionnaire. J Hand Surg Am. 2005, 30: 1074, e1–10.

Klum M, Wolf MB, Hahn P, Leclère FM, Bruckner T, Unglaub F. Normative data on wrist function. J Hand Surg Am. 2012, 37: 2050–60.

MacDermid JC. Outcome evaluation in patients with elbow pathology: issues in instrument development and evaluation. J Hand Ther. 2001, 14: 105–14.

MacDermid JC, Richards RS, Donner A, Bellamy N, Roth JH. Responsiveness of the short form-36, disability of the arm, shoulder, and hand questionnaire, patient-rated wrist evaluation, and physical impairment measurements in evaluating recovery after a distal radius fracture. J Hand Ther. 2000, 25: 330–40.

MacDermid JC, Turgeon T, Richards RS, Beadle M, Roth JH. Patient rating of wrist pain and disability: a reliable and valid measurement tool. J Orthop Trauma. 1998, 12: 577–86.

Snyder CF, Aaronson NK, Choucair AK et al. Implementing patient-reported outcomes assessment in clinical practice: a review of the options and considerations. Qual Life Res. 2012, 21: 1305–14.

Chapter 6 - CONCLUSIONS

Discussion

It is an important goal to differentiate normal, age-related changes in joint function from those changes associated with symptomatic pathology; surgical intervention; or implant wear, fatigue, or failure. A significant step toward this goal is the establishment of a joint-specific PROM reference database for individuals without joint disease, so that clinicians can effectively evaluate the efficacy of treatments in orthopaedic patients on a longitudinal basis.

This thesis has investigated and reported the influence on PROM values of age, gender, ethnicity, handedness (where applicable), history of previous injury/pathology/surgery and coincident adjacent active joint pathology, in two different countries. No other research exists comparing electronically-collected PROMs for multiple body regions across two continents. This research represents the only normative data of several PROMs investigated [11, 17, 49, 56]. It also represents the largest database of combined normative PROM values reported in the literature (n=2360). No other study has compared normal scores in international cohorts or collected them remotely and electronically via the same research database.

This thesis involved the investigation of multiple PROMs, including both subjective and combined subjective/objective PROMs, the latter requiring additional resources to collect. The PROMs investigated cover the four major joints for which the majority of literature exists (hip [13, 23], knee [10, 18], shoulder [12, 14, 15, 47-50] and wrist/hand PROM [11, 12]). The collected data has established a database of PROM values for individuals who identify themselves as asymptomatic for the body parts under investigation. These pooled values can be used as control cohorts for future studies investigating pathological cohorts.

This thesis adopted carefully chosen inclusion and exclusion criteria in an effort to establish asymptomatic cohorts (individuals who identified themselves as ‘normal’), for the PROMs being investigated. Previous studies reporting the collection of “normative values” have not made attempts to exclude subjects with pathology to the same extent as the present study [48, 54-56, 65, 66, 72-75, 78]. The inclusion of participants of different ages, genders and ethnicity, as well as not excluding participants with a history of previous non-arthroplasty surgery and coincident adjacent active joint pathology, represents a significant addition to the reported PROM literature. These important factors may well have been present in pathological cohorts presented in the literature but not necessarily recognised and/or recorded.

This thesis reports that differences in age, gender, handedness (when applicable), ethnicity, history of injury/pathology/surgery and nationality, need to be considered when using PROMs to assess patient outcomes. Often, an inverse relationship is observed between age and clinical score [53, 55, 64]. 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 a reported functional score independent of pathology[64]. The large variation in PROM reports in individuals over the age of 80 years suggests that there may be other determinants of health and function that influence the subjective score reported, and possibly the accuracy of asymptomatic "normal" PROM values in this age older group.

A detailed discussion of the individual study findings and a comparison with the current literature is presented in the Discussion of each manuscript Chapter.

This thesis supports the conclusion that an asymptomatic "normal" control group should ideally be sourced from the same country of origin and be age- and gender-matched. It could include either subjective-only or combined subjective-objective PROM inputs without compromising the outcome data, depending on the investigators' preference or available resources.

Other considerations such as *a history of active adjacent joint pathology* or *a history of inactive pathology* in the joint under investigation, will likely have varying influence on the reported PROM value. The choice of a PROM that includes questions that are more joint-specific may help avoid introducing bias attributable to these confounders. An appreciation of the influence that coincident adjacent or contralateral joint pathology or *a history of an inactive pathology* may have on reported pathological PROM values will help future investigators in their interpretation of pathological patient cohorts.

Certainly, when considering future studies using an electronic methodology and a pre-established control group, all these variables should be considered when preparing clinical publications with outcome data over a 2 year follow-up period. This might well enhance the accuracy of future studies and improve our evidence-based practice of orthopaedic surgery.

Limitations

This thesis identified several limitations. One limitation is the subjective interpretation of 'normal' or 'asymptomatic'. Indeed, achieving a 'normal' comparative cohort relies on subjective individual reports of pain, movement and function. As radiographs were not performed to rule out asymptomatic, incidental degenerative disease, it is possible that some individuals with asymptomatic pathology were included in each cohort.

When considering any randomly collected cohort, it is assumed that a small percentage of the general population will have asymptomatic, incidental undiagnosed pathology. The cohorts in this thesis are therefore likely representative of this natural variation in the general population.

In addition, participants with a history of pathology treated by non-arthroplasty/arthrodesis surgery >more than 3 years ago and no longer experiencing symptoms were included. It is possible that some of these participants continued to suffer from some slight impairment despite their subjective perception that their recovery was complete and their joint was no longer considered to be a problem. When considering future studies, it is recommended that these patients not be removed, as the incidence of participants with *a history of pathology treated by non-arthroplasty/arthrodesis surgery* >*3yrs*, increases with increasing age. If these individuals are removed, this would introduce a selection bias; especially as these individuals had identified themselves as *asymptomatic* prior to completing their PROMs.

As with any observational study, there is the potential for selection bias. Although these studies collected individuals randomly, no specific randomised method of recruitment was employed. The primary benefits of randomization 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 randomization method of participant identification was employed. Another potential source of procedural bias was the use of electronic questionnaires, as some elderly participants may have declined to be involved due to the use of technology.

Despite these limitations, this thesis presents a considered methodology, conceptualized in consultation with epidemiologists, to establish as close to 'asymptomatic' or 'normal' cohorts as practically possible within the constraints of ethics, budget, time, expertise and resources. The collated database will facilitate the establishment of control cohorts in the future when considering pathological cohorts for comparison.

Future Directions

Several other studies have been performed in conjunction with the work completed as part of this thesis. Although not formally included as part of this thesis, these publications complement the present study's aims and will augment planned future studies (Appendix 1 and 2).

This thesis identified shoulder range of motion (ROM) as a potential variable that could be assessed remotely, in a similar manner to subjective-only PROMs. However, accurate measurement of shoulder range of motion (ROM) via a reproducible, reliable and validated electronic method had not been established. An electronic method has the potential for remote assessment and transfer of data. This thesis identified a knowledge deficiency in the assessment of shoulder ROM devices currently in use. The aim of the additional study was to establish the reliability and validity of different smartphone applications (apps) in assessing pathologic shoulder ROM and to determine whether differences in recorded ROM measurements affect calculated shoulder scores. It was hypothesized that there is no difference between shoulder ROM assessment methods and calculated shoulder scores. The study involved the recruitment of 75 patients with a variety of shoulder pathologies. Participants were assessed using a smartphone inclinometer, a smartphone virtual goniometer, a standard goniometer, and clinicians' visual estimation. Shoulder strength was assessed and CSS and UCLA shoulder scores were calculated. The outcomes of this study are presented in Appendix 1.

This thesis identified the importance of comparing a pathological patient cohort with an international, age- and gender-matched, asymptomatic "normal" control cohort. Investigating this knowledge deficiency, would help determine if an asymptomatic "normal" control group should be sourced from the same country of origin as the derived study. The aim of the additional study was to determine whether an electronic, multicenter data collection system could be used to establish normal population reference values for the Hip Disability and Osteoarthritis Outcome Score (HOOS) and the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC). The secondary aim was to investigate differences in asymptomatic HOOS and WOMAC values reported in 2 geographically distinct English-speaking countries and compare these with a symptomatic arthritic patient cohort. The study involved the recruitment of 552 total participants, including asymptomatic "normal" participants and end-stage, radiographically confirmed osteoarthritis planned for total hip arthroplasty surgery. Participants completed electronic HOOS and WOMAC PROMs using similar methods and analyses to those presented in the Chapters included in this thesis. The outcomes of this study are presented in Appendix 2.

Several studies are planned to follow-on from the research completed as part of this thesis. The Author is currently contributing to the Australian Orthopaedic Association National Joint Replacement Registry

(AOANJRR) Patient Reported Outcome Measures (PROMs) Pilot Project. The normative data contained in this thesis has been offered to Professor Stephen Graves as a control group for comparison for the AOANJRR PROMs Pilot Project.

Long-term, the Author has the intention of making the data contained in this thesis available as an electronic control group to the AOANJRR, for use by all Australasian orthopaedic surgeons and their patients. The establishment of a normal de-identified control group within an electronic database will aid in improving the robustness and quality of future studies and assessing patient outcomes.

An international database has the potential to collect de-identified data from different regions remotely and use this data to complement multi-centre, international clinical studies that are robust and capable of rapidly advancing our understanding of clinical outcomes. Overall, this will significantly enhance our ability to collect reliable, validated data and enhance our practice of evidence-based medicine and surgery.

Summary

The work described in this thesis has addressed the perceived deficiencies in the knowledge of PROM values in normal, asymptomatic individuals. The thesis encompasses studies of commonly employed validated PROMs used in the clinical assessment of conditions of the hip, knee, shoulder and wrist/hand PROMs.

The problem areas identified, which were addressed, included (i) the establishment of asymptomatic population reference values for commonly used hip, knee, shoulder and wrist/hand PROMs (ii) the identification of factors other than pathology that influence reported PROM values and, (iii) the establishment of an electronic data collection system for future comparisons with pathological patient cohorts.

To address these deficiencies, PROM data was collected electronically and remotely across two continents, from individuals who identified themselves as asymptomatic in the joint under question. Factors such as age, gender, handedness (when applicable), history of pathology, ethnicity and nationality, were associated with variations in reported PROM values reported. Further research has commenced to examine how these variables are influenced in a pathological cohort.

The work described in this thesis increases the knowledge of PROM interpretation, application and investigation. It lies the groundwork for future studies through the establishment of a database of 2360 asymptomatic “normal” PROM population reference values, which can be used in the establishment of control group(s) for comparison with pathological and surgical cohorts. It further advances the role of clinical follow-up in the treatment of orthopaedic conditions and our practice of evidence-based medicine.

APPENDIX 1

SMARTPHONE APPLICATIONS FOR THE EVALUATION OF PATHOLOGICAL SHOULDER RANGE OF MOTION AND SHOULDER SCORES – A COMPARATIVE STUDY

As published in Journal of Shoulder & Elbow Surgery. 2018; 2(1): 109-114.

JSES Open Access 2 (2018) 109–114

Contents lists available at ScienceDirect

JSES Open Access

journal homepage: www.elsevier.com/locate/jses

Smartphone applications for the evaluation of pathologic shoulder range of motion and shoulder scores—a comparative study

Kevyn Mejia-Hernandez, BMBS, MSc [a,b], Angela Chang, MBBS [a,b], Nathan Eardley-Harris, MBBS [b], Ruurd Jaarsma, MD, PhD, FRACS (ORTHO) [b], Tiffany K. Gill, BAppSc (Physio), MAppSc (Physio), CertHlthEc, PGradD [a], James M. McLean, MS, MBBS, FRACS [a,*]

[a] *Discipline of Orthopaedics and Trauma, University of Adelaide, Adelaide, SA, Australia*
[b] *Discipline of Orthopaedics and Trauma, Flinders Medical Centre, Adelaide, SA, Australia*

ARTICLE INFO

Keywords:
Shoulder range of motion
Inclinometer
Goniometer
Smartphone application
Constant-Murley score
UCLA score

Level of evidence: Level III, Diagnostic Study

Hypothesis and background: Accurate measurement of range of motion (ROM) is important in evaluating a pathologic shoulder and calculating shoulder scores. The aim of this study was to establish the reliability and validity of different smartphone applications (apps) in assessing pathologic shoulder ROM and to determine whether differences in recorded ROM measurements affect calculated shoulder scores. The authors hypothesized that there is no difference between shoulder ROM assessment methods and calculated shoulder scores.

Methods: In this nonrandomized controlled clinical trial, ROM of 75 participants with a history of shoulder disease (21 women, 54 men) was assessed using a smartphone inclinometer and virtual goniometer, a standard goniometer, and clinicians' visual estimation. Shoulder strength was assessed, and Constant-Murley (CM) and University of California–Los Angeles (UCLA) shoulder scores were calculated.

Results: Independent of diagnosis or operation, all cases (except for passive glenohumeral abduction of unstable shoulders) showed excellent intraclass correlation coefficients (>0.84). Interobserver reliability was excellent for all ROM measures (intraclass correlation coefficient > 0.97). All modalities had excellent agreement to values attained with the universal goniometer. There were no differences for the calculated CM or UCLA scores between the modalities employed to measure ROM.

Conclusions: A smartphone inclinometer or virtual goniometer is comparable to other clinical methods of measuring pathologic shoulder ROM. Clinicians can employ smartphone applications with confidence to measure shoulder ROM and to calculate UCLA and CM scores. The apps are also available to patients and may be a useful adjunct to physiotherapy, especially in cases of limited access to health care services.

Accurate and reliable measurement of shoulder range of motion (ROM) is integral to the physical examination and functional evaluation of a pathologic shoulder. A universal goniometer (UG) is considered the "gold standard" for measuring shoulder ROM[1]; however, visual estimation is common in clinical practice as it is more time efficient, and a goniometer is often not available.[2,26,28] Other methods for measuring shoulder ROM include digital inclinometry, digital motion capture, and high-speed cinematography, but these require expensive, specialized equipment with limited availability.[9,10,13,16,24,28]

Smartphone applications (apps) have recently been proposed as an alternative method of measuring pathologic shoulder ROM.[15,21,22,26,29] Apps rely on an internal smartphone inclinometer[26] or a photographic virtual goniometer[21] to measure ROM. Several studies have demonstrated joint ROM measured with apps to be reliable and accurate compared with traditional methods,[7,14,15,19,21,23,25,26,29] but studies performed on the shoulder were limited by inclusion of only participants with no joint disease (for whom they have the most potential clinical application). In addition, no shoulder study considered the impact that ROM variability may have on shoulder scores with an objective ROM component.

The Constant-Murley (CM) score[5] and the University of California–Los Angeles[11] (UCLA) shoulder score are commonly used shoulder assessment tools that evaluate level of function and efficacy of

The Human Research Ethics Committee, The Queen Elizabeth Hospital, Basil Heizel Institute DX465101, approved this study: HREC/14/TQEHLMH/239.

* Corresponding author: James M. McLean, MS, MBBS, FRACS, Discipline of Orthopaedics and Trauma, University of Adelaide, North Terrace, Adelaide, SA 5005, Australia.
E-mail address: james.mc_lean@adelaide.edu.au (J.M. McLean).

https://doi.org/10.1016/j.jses.2017.10.001

surgical interventions and observe clinical change over time. These shoulder patient-reported outcome measures have the advantage over other scoring systems of including subjective patient-derived inputs as well as objective clinician-derived inputs, allowing a more balanced interpretation of shoulder function.[5]

Shoulder scores enhance communication during the physician-patient consultation[6,8] and aid in clinical decision-making. As shoulder ROM is an important component of these scores, measurements must be accurate and reliable for the scores to be of clinical use.

Apps have been shown to be accurate and reliable in measuring ROM in normal shoulders; however, their use in the pathologic shoulder is yet to be assessed. The authors hypothesized that smartphone ROM apps will provide accurate and reliable measurements when tested on pathologic shoulders.

Materials and methods

Patients were recruited from the outpatient department of 2 tertiary orthopedic units between February 2015 and February 2016. Inclusion criteria were being English speaking, older than 18 years, and willing to provide informed consent and having a documented current shoulder disease. Patients were excluded if they had cognitive impairment or were unable to follow the assessor's instructions. In cases of bilateral shoulder disease, both shoulders were assessed independently and included.

One iPhone (Model 5S) was used in the study and the software not updated during data collection. Two iPhone apps were used to measure shoulder range of movement (ROM): GetMyROM (version 1.0.3; Interactive Medical Productions, Hampton, NH, USA), an inclinometry-based app (Fig. 1, *A*); and DrGoniometer (version 1.2; CDM S.r.L, Milano, Italy), a photo capture–based application (Fig. 1, *B*). Visual ROM estimates were recorded for each subject, as were measurements made using a standard, manual goniometer as a control. A questionnaire recorded the subjective and functional questions of the UCLA and CM shoulder scores.

All participants were assessed with exposed shoulders. Two medical practitioner observers with experience in musculoskeletal disease collected the data independently with an assistant. Participants initially sat upright and straight on a fixed chair to stabilize the spine. In this position, the following measurements were observed: active forward flexion (Fig. 2, *A*), total abduction, active glenohumeral abduction (Fig. 2, *B*), and passive glenohumeral abduction. To assess glenohumeral joint abduction, the participants were asked to abduct the arm while the examiner stabilized the scapular. Commencement of scapula rotation was used to determine the limit of glenohumeral joint movement.

Rotation of the shoulder was measured with participants supine on a standard examination table. The shoulder was positioned in 90° of abduction with 90° of flexion at the elbow. With the forearm in neutral rotation and the proximal two-thirds of the humerus supported by the table, measurements were taken for active (Fig. 2, *C*) and passive external rotation and active (Fig. 2, *D*) and passive internal rotation. If shoulder disease prevented the participant from abducting the shoulder to 90°, supine external rotation was measured with the elbow in contact with the side of the body (0° abduction), and internal rotation measures were not recorded. With all ROM tests, care was taken to avoid compensatory movements, such as elbow extension or scapular elevation, and if these were observed, the measurement was repeated.

Shoulder ROM was first assessed using the smartphone inclinometer attached to the participant with a DualFit Armband (Belkin; Playa Vista, CA, USA). The armband was attached to the distal portion of the humerus for seated movements, then repositioned to the wrist for measurements performed with the participant supine. The inclinometer was positioned with the screen facing away from the observer. The assistant read and recorded the ROM value with the observer blinded to the reading. Next, the observer captured

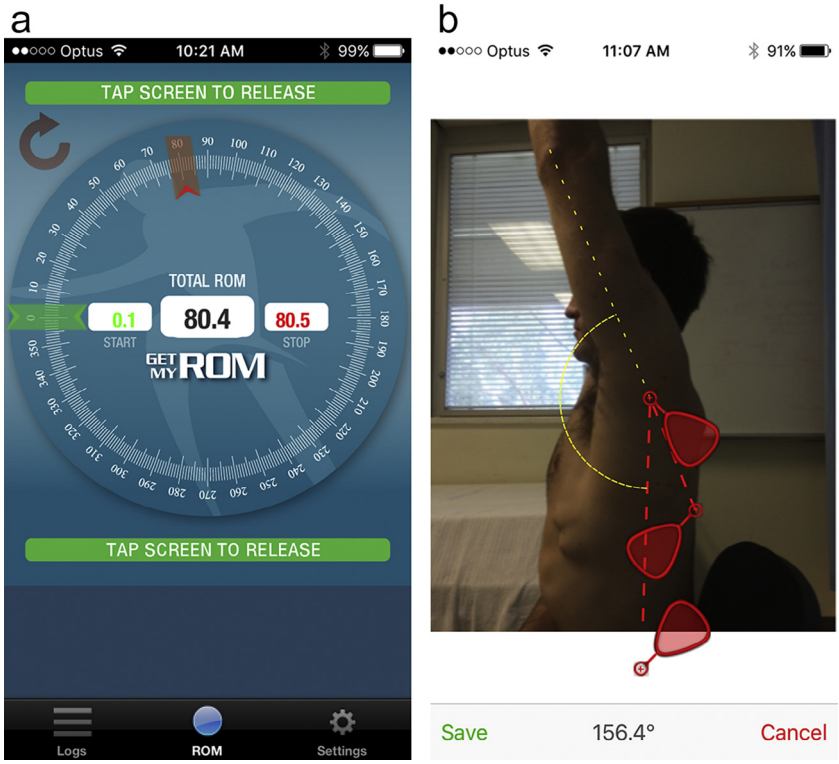

Figure 1 The iPhone applications. (**A**) GetMyROM. (**B**) DrGoniometer.

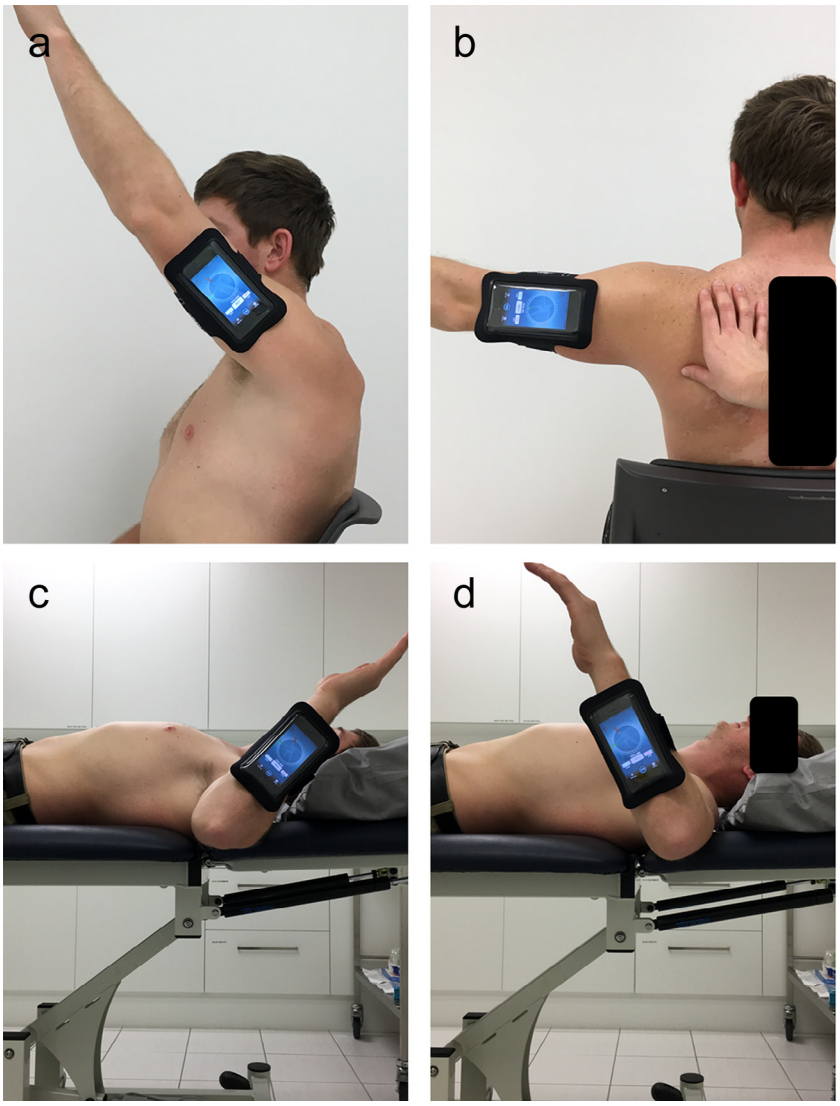

Figure 2 Measurements of (**A**) forward flexion, (**B**) glenohumeral abduction, (**C**) external rotation, and (**D**) internal rotation.

photographs of the participant with the shoulder positioned at the limits of ROM, to be analyzed at a later date using DrGoniometer. Third, the observer gave a visual estimate of shoulder ROM and last measured the ROM using the standard goniometer.

A myometer (Isometer; Innovative Design Orthopaedics Ltd., London, UK) was employed to measure shoulder strength using the method described by Constant et al.[5] Participants stood with their feet shoulder-width apart and the arm held at 90° of abduction in the scapular plane. The forearm was placed in full pronation and shoulder internal rotation. A looped arm strap was placed 2 cm proximal to the wrist. With the elbow straight at all times, participants were instructed to lift up with maximal strength. The measurement was repeated 2 more times and the highest value used to calculate the CM score.[5,29] Participants who were not able to establish the desired degrees of abduction or experienced pain when completing this part of the assessment were assigned a strength score of 0 and were not subjected to further strength measurement. Strength measurements were used in conjunction with the questionnaires to calculate the UCLA and CM shoulder scores.

To assess the reliability of the 4 methods, all measurements were completed by 2 independent observers. An intraclass correlation coefficient (ICC) was calculated using SAS 9.3 (SAS Institute, Cary, NC, USA). An ICC was calculated for each ROM test, interpreted as follows: 0.00-0.40, poor correlation; 0.41-0.59, fair correlation; 0.60-0.74, good correlation; and 0.75-1.00, excellent correlation.[4]

A Bland-Altman analysis was used to assess the level of agreement between the ROM measurement modalities. A 95% limit of agreement was defined as ±1.96 standard deviations around the mean difference from measurements obtained by the UG, thus producing a lower and upper level of agreement.

A negative binomial generalized estimating equation was used to account for clustering on random effects of the subjects and raters. Assumptions of linear regression were not upheld. Modeling was performed for the CM scores and ROM as well as for UCLA scores and ROM.

Results

From February 2015 to February 2016, 75 patients were recruited from the outpatient department of 2 tertiary orthopedic units (21 women, 54 men). The average age was 46 years (range, 24-94 years). Fourteen patients had bilateral shoulder disease.

The underlying shoulder diseases in descending order of frequency included shoulder instability (n = 23), degenerative changes/

Table I
Diagnoses and operations of participants (N = 75)

Diagnosis	Operation
Inflammation (n = 9)	No operation (n = 58)
SA bursitis ± impingement	Arthroscopic surgery (n = 6)
Scapulothoracic bursitis/snapping scapula	Rotator cuff repair
Fractures (n = 4)	Labral stabilization
Proximal humerus	Other therapeutic procedure
Clavicle	Open nonarthroplasty procedures (n = 10)
Degenerative/arthritic (n = 9)	Latarjet/Bristow
ACJ osteoarthritis	ACJ lateral clavicle excision
SCJ osteoarthritis	Rotator cuff repair
GHJ osteoarthritis	Open washout of shoulder joint
Septic arthritis	Total shoulder arthroplasty (n = 1)
Instability (n = 23)	Reverse
Traumatic (including dislocations)	
ACJ dislocation/subluxation	
Soft tissue (n = 30)	
Proximal biceps tendon tear (long head)	
Rotator cuff tear	
SLAP biceps tear	
Supraspinatus tendinopathy	

SA, subacromial; *ACJ*, acromioclavicular joint; *SCJ*, sternoclavicular joint; *GHJ*, glenohumeral joint; *SLAP*, superior labral tear from anterior to posterior.

arthritis (n = 9), inflammation (n = 9), prior fracture (n = 4), and soft tissue disease (n = 30). Diagnoses and surgical interventions are presented in Table I.

In all cases for the group diagnoses (except for passive glenohumeral abduction of the instability group), ICC values were classified as having excellent agreement (>0.84). In regard to surgical intervention, the ICC value for each ROM test was >0.95.

The ICC values are presented in Table II. All values were >0.97, indicating excellent agreement between the 4 methods of measurement for each ROM. Each ROM variable showed a left-skewed distribution.

Table II
Interobserver reliability for each range of motion (ROM) measurement method

ROM test	Method	ICC	Lower 95% limit	Upper 95% limit
Forward flexion	Clinician	1.00	0.99	1.00
	Goniometer	0.99	0.99	1.00
	GetMyROM	0.99	0.98	0.99
	DrGoniometer	1.00	0.99	1.00
Total active abduction	Clinician	1.00	0.99	1.00
	Goniometer	0.99	0.99	1.00
	GetMyROM	0.99	0.99	1.00
	DrGoniometer	0.99	0.99	1.00
Active glenohumeral abduction	Clinician	0.99	0.99	1.00
	Goniometer	0.98	0.97	0.99
	GetMyROM	0.98	0.97	0.99
	DrGoniometer	0.98	0.97	0.99
Passive glenohumeral abduction	Clinician	0.99	0.99	1.00
	Goniometer	0.98	0.99	0.99
	GetMyROM	0.97	0.95	0.99
	DrGoniometer	0.98	0.96	0.99
Active internal rotation	Clinician	0.99	0.99	0.99
	Goniometer	0.99	0.99	0.99
	GetMyROM	0.98	0.97	0.99
	DrGoniometer	0.99	0.99	1.00
Passive internal rotation	Clinician	0.99	0.98	0.99
	Goniometer	0.99	0.99	0.99
	GetMyROM	0.98	0.96	0.98
	DrGoniometer	1.00	0.99	1.00
Active external rotation	Clinician	0.99	0.98	0.99
	Goniometer	0.99	0.98	0.99
	GetMyROM	0.99	0.98	0.99
	DrGoniometer	1.00	0.99	1.00
Passive external rotation	Clinician	0.99	0.99	1.00
	Goniometer	1.00	0.99	1.00
	GetMyROM	0.99	0.98	0.99
	DrGoniometer	0.99	0.99	1.00

ICC, intraclass correlation coefficient.

Table III
Bland-Altman plots comparing measurement modalities to the "gold standard" goniometer

Comparison measurement mean to universal goniometer	Comparative measurement technique	Mean difference	Lower 95% limit	Upper 95% limit
Forward flexion	Clinician	1.69	–5.17	8.55
	GetMyROM	–0.76	–9.64	8.11
	DrGoniometer	–0.56	–9.63	8.52
Total active abduction	Clinician	0.37	–6.46	7.20
	GetMyROM	0.47	–7.87	8.81
	DrGoniometer	0.81	–7.73	9.35
Active glenohumeral abduction	Clinician	–0.18	–2.90	2.54
	GetMyROM	–0.19	–4.71	4.32
	DrGoniometer	–0.41	–6.93	6.12
Passive glenohumeral abduction	Clinician	–0.07	–2.04	1.90
	GetMyROM	–0.38	–4.02	3.25
	DrGoniometer	–0.01	–2.32	2.30
Active internal rotation	Clinician	–0.29	–5.51	4.93
	GetMyROM	0.51	–7.11	8.14
	DrGoniometer	–1.29	–10.00	7.43
Passive internal rotation	Clinician	0.00	–4.66	4.66
	GetMyROM	0.55	–5.04	6.13
	DrGoniometer	–1.41	–9.83	7.01
Active external rotation	Clinician	0.01	–5.02	5.05
	GetMyROM	–0.08	–8.32	8.17
	DrGoniometer	0.20	–8.23	8.63
Passive external rotation	Clinician	0.42	–5.18	6.02
	GetMyROM	0.40	–7.58	8.37
	DrGoniometer	0.13	–7.65	7.91

The results of Bland-Altman plots, comparing the other 3 methods of measurement with the measurements obtained by the UG, are presented in Table III. These values are within a narrow range and indicate generally superior agreement, although active and passive glenohumeral abduction had the narrowest limit range (ie, more agreement between the methods of measurement and UG), and forward flexion resulted in a larger range of values.

Nineteen participants (25.3%) were not able to complete strength assessments as a result of pain (n = 15) or loss of ROM (n = 4). These participants were assigned a strength score of 0 in calculating UCLA and CM scores. When the CM and UCLA scores were calculated for each method of assessing ROM, all methods revealed identical mean CM scores (74; range, 6-100) and mean UCLA scores (29; range, 7-35).

Discussion

This study demonstrated that the ROM measurements in pathologic shoulders are consistent in comparing a UG with visual estimates and 2 different smartphone ROM apps. It also demonstrated that there is no difference in the calculated shoulder UCLA and CM scores using the 4 different methods, suggesting that newer technologies that use smartphone applications may be a useful tool in the clinical setting.

Previous studies[3,12,17,30] have assessed the accuracy and validity of smartphone apps in measuring joint ROM; however, those focused on the shoulder were limited by assessing only normal shoulders in healthy, young individuals.[15,18,21] In contrast, this study assessed patients of varying ages with a spectrum of shoulder diseases, thus providing a relevant clinical context and a broader spectrum of assessable shoulder ranges over which to compare the assessment modalities. Previous studies were also limited to the assessment of active shoulder motion only.[15,18,21] In addition to examining both active and passive ROM, this study also assessed a more complete set of shoulder movements, including internal and external rotation, abduction, and forward flexion. This study also assessed the

clinically relevant movement of glenohumeral joint abduction, an important sign that can assist in differentiating subacromial and subdeltoid adhesions from adhesive capsulitis in pathologic and postoperative patients. To the authors' knowledge, this movement has not previously been investigated using different ROM assessment modalities.

This study is the first to compare 4 different shoulder ROM assessment tools in a pathologic patient cohort. The 4 assessment modalities used (the clinician's visual estimation, inclinometer-based smartphone application [GetMyROM], photograph-based [DrGoniometer] smartphone application, and UG) demonstrated excellent agreement. This finding was similar to that of Werner et al, who assessed postsurgical shoulder ROM in patients having undergone total shoulder replacement surgery and reported excellent correlation when using an inclinometer Smartphone application in isolation.[29] This study is also one of the few studies that assessed an older population cohort (mean age, 46 years), in various planes of shoulder movement, with various shoulder diseases and diagnoses. The only other study to examine inclinometer- and photograph-based shoulder assessment tools relied exclusively on measurements of external rotation performed on young, healthy subjects (mean age, 26.4 ± 7.6 years).[21]

Certain advantages exist in adopting these new technologies. Inclinometry-based applications allow fast, reliable measurements of shoulder ROM and are widely available and cost-effective, given the prevalence of smartphone ownership in the general population. They are available not only to physicians but also to allied health professionals and patients. As well as being used in a clinical setting, the inclinometry-based apps allow patient self-measurement,[12] providing real-time feedback for exercise completed at home. This may be of particular benefit to those with limited access to health care because of rural location or disability, for whom some assessment may be performed by a combination of telephone, tele-link, or secured e-mail.

Photograph-based applications, whereby clinicians make measurements in a delayed fashion (post-production and independent of the patient's location), allow images to be printed and filed in patient notes for comparison during subsequent visits. Like inclinometry-based apps, photograph-based applications allow accurate ROM assessment when a goniometer is not available or when a face-to-face interaction with a health professional is not immediately available. Moreover, the physician-patient interaction can potentially be enhanced by demonstrating the patient's progress in ROM over time.[12,20]

Whereas visual assessment of ROM may require experience to give an accurate estimation, a recent study[29] reported that the skill level of medical assessor does not influence the ROM assessment with use of smartphone applications (ie, student vs. medical clinician). This is important, as clinicians and allied health practitioners who do not have exclusively musculoskeletal practices may employ these smartphone applications with the confidence that they will produce consistent results. Non–musculoskeletal-focused clinicians are less likely to have experience with visual estimation of shoulder ROM, and consequently the measurements obtained may be affected.

The UCLA and CM scores are important adjuncts to the management of patients with shoulder disease and can be used to assess for change in function over time. Both have subjective and objective components that allow a more balanced interpretation of a patient's true shoulder function. This is the first study to compare smartphone applications for measuring shoulder ROM in a pathologic patient cohort and subsequently using these measurements to calculate and to compare UCLA and CM shoulder scores. In this study, both the UCLA and CM shoulder scores were similar in comparing various modalities of ROM measurement. These findings suggest that these smartphone applications can be used with confidence to calculate UCLA and CM shoulder scores in patients with a spectrum of shoulder diseases.

Limitations exist in using inclinometer- and photograph-based smartphone applications for shoulder ROM measurement. Inclinometer applications require the mobile device to be in contact with the patient. This is best achieved with an instrument that physically holds the phone to the patient (such as an armband).[26] Hygiene issues may be raised when an armband is used repeatedly and suggest the need for disposable armbands.[1,27] The position of the measurement device is important to achieve consistent results. It was our experience that the armband sometimes required adjustment (eg, rotated around the longitudinal axis of the humerus) to attain the appropriate measure. Vigilance of the assessor was required to adjust the armband if it slipped or loosened after initial application.

Photograph-based smartphone applications had similar considerations, especially when the examiner was taking the picture and in positioning the virtual goniometer on anatomic landmarks to measure the angles under investigation.[12] Errors could be made if the picture was taken short of maximal ROM or if the photograph was mistimed (and not picked up). Care needed to be taken in measuring the angles under investigation in the photographs, especially in placing markers on the desired anatomic landmarks.

There are limitations to this study. The 2 observers were highly trained and had many years of experience in shoulder examination. This may have influenced the visual estimation results, which may not be reproducible for clinicians with less experience. Care should be taken in interpreting this result, as visual estimation requires a certain level of expertise and practice, which may take several years to achieve. This modality has the most inherent variability and is not recommended for routine use (especially for clinicians who have not self-evaluated their assessments). Ongoing self-evaluation is required to confirm that this technique is valid for each clinician.

Another limitation was the sample size, which was not randomized and was composed of more men than women. The average age of the sample was skewed to the elderly (likely representative of the larger proportion of disease found in the older age groups). A broader and larger sample would allow a more in-depth analysis by gender and age subsets. Future studies could incorporate examination of dynamic movements in addition to static assessments.

Conclusion

Technologic advances offer the opportunity to adopt new tools that can improve patient assessments and follow-up and ultimately lead to improved clinical outcomes. Smartphone app use is a widely available, cost-effective method to assist clinicians in accurately measuring joint range of movement, including that of the shoulder, knee, spine, elbow, and ankle.[3,12,17,30] This study demonstrates that shoulder ROM can be reproducibly measured using 4 independent methods. Smartphones can be used with confidence by clinicians to provide a reliable, reproducible, practical, and inexpensive way of assessing shoulder ROM.

Acknowledgments

The authors thank Suzanne Edwards, Statistician, Data Management and Analysis Centre at the University of Adelaide, for her help with the statistical analysis of the collected data.

Disclaimer

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.

The authors, their immediate families, and any research foundations with which they are affiliated have not received any financial payments or other benefits from any commercial entity related to the subject of this article.

References

1. Albrecht UV, von Jan U, Sedlacek L, Groos S, Suerbaum S, Vonberg RP. Standardized, App-based disinfection of iPads in a clinical and nonclinical setting: comparative analysis. J Med Internet Res 2013;15:e176. http://dx.doi.org/10.2196/jmir.2643
2. Awan R, Smith J, Boon AJ. Measuring shoulder internal rotation range of motion: a comparison of 3 techniques. Arch Phys Med Rehabil 2002;83:1229-34. http://dx.doi.org/10.1053/apmr.2002.34815
3. Blonna D, Zarkadas PC, Fitzsimmons JS, O'Driscoll SW. Validation of a photography-based goniometry method for measuring joint range of motion. J Shoulder Elbow Surg 2012;21:29-35. http://dx.doi.org/10.1016/j.jse.2011.06.018
4. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychol Assess 1994;6:284-90.
5. Constant CR, Gerber C, Emery RJ, Sojbjerg JO, Gohlke F, Boileau P. A review of the Constant score: modifications and guidelines for its use. J Shoulder Elbow Surg 2008;17:355-61. http://dx.doi.org/10.1016/j.jse.2007.06.022
6. Coons SJ, Gwaltney CJ, Hays RD, Lundy JJ, Sloan JA, Revicki DA, et al. Recommendations on evidence needed to support measurement equivalence between electronic and paper-based patient-reported outcome (PRO) measures: ISPOR ePRO Good Research Practices Task Force report. Value Health 2009;12:419-29. http://dx.doi.org/10.1111/j.1524-4733.2008.00470.x
7. Cuesta-Vargas AI, Roldan-Jimenez C. Validity and reliability of arm abduction angle measured on smartphone: a cross-sectional study. BMC Musculoskelet Disord 2016;17:93. http://dx.doi.org/10.1186/s12891-016-0957-3
8. Dawson J, Fitzpatrick R, Carr A. Questionnaire on the perceptions of patients about shoulder surgery. J Bone Joint Surg Br 1996;78:593-600.
9. De Winter AF, Heemskerk MA, Terwee CB, Jans MP, Deville W, Schaardenburg DJ, et al. Inter-observer reproducibility of measurements of range of motion in patients with shoulder pain using a digital inclinometer. BMC Musculoskelet Disord 2004;5:18. http://dx.doi.org/10.1186/1471-2474-5-18
10. El-Zayat BF, Efe T, Heidrich A, Wolf U, Timmesfeld N, Heyse TJ, et al. Objective assessment of shoulder mobility with a new 3D gyroscope—a validation study. BMC Musculoskelet Disord 2011;12:168. http://dx.doi.org/10.1186/1471-2474-12-168
11. Ellman H, Hanker G, Bayer M. Repair of the rotator cuff. End-result study of factors influencing reconstruction. J Bone Joint Surg Am 1986;68:1136-44.
12. Ferriero G, Vercelli S, Sartorio F, Munoz Lasa S, Ilieva E, Brigatti E, et al. Reliability of a smartphone-based goniometer for knee joint goniometry. Int J Rehabil Res 2013;36:146-51. http://dx.doi.org/10.1097/MRR.0b013e32835b8269
13. Hayes K, Walton JR, Szomor ZR, Murrell GA. Reliability of five methods for assessing shoulder range of motion. Aust J Physiother 2001;47:289-94.
14. Jenny JY. Measurement of the knee flexion angle with a Smartphone-application is precise and accurate. J Arthroplasty 2013;28:784-7. http://dx.doi.org/10.1016/j.arth.2012.11.013
15. Johnson LB, Sumner S, Duong T, Yan P, Bajcsy R, Abresch RT, et al. Validity and reliability of smartphone magnetometer-based goniometer evaluation of shoulder abduction—a pilot study. Man Ther 2015;20:777-82. http://dx.doi.org/10.1016/j.math.2015.03.004
16. Jordan K, Dziedzic K, Jones PW, Ong BN, Dawes PT. The reliability of the three-dimensional FASTRAK measurement system in measuring cervical spine and shoulder range of motion in healthy subjects. Rheumatology (Oxford) 2000;39:382-8.
17. Kolber MJ, Pizzini M, Robinson A, Yanez D, Hanney WJ. The reliability and concurrent validity of measurements used to quantify lumbar spine mobility: an analysis of an iphone application and gravity based inclinometry. Int J Sports Phys Ther 2013;8:129-37.
18. Lim JY, Kim TH, Lee JS. Reliability of measuring the passive range of shoulder horizontal adduction using a smartphone in the supine versus the side-lying position. J Phys Ther Sci 2015;27:3119-22. http://dx.doi.org/10.1589/jpts.27.3119
19. Milanese S, Gordon S, Buettner P, Flavell C, Ruston S, Coe D, et al. Reliability and concurrent validity of knee angle measurement: smart phone app versus universal goniometer used by experienced and novice clinicians. Man Ther 2014;19:569-74. http://dx.doi.org/10.1016/j.math.2014.05.009
20. Milani P, Coccetta CA, Rabini A, Sciarra T, Massazza G, Ferriero G. Mobile smartphone applications for body position measurement in rehabilitation: a review of goniometric tools. PM R 2014;6:1038-43. http://dx.doi.org/10.1016/j.pmrj.2014.05.003
21. Mitchell K, Gutierrez SB, Sutton S, Morton S, Morgenthaler A. Reliability and validity of goniometric iPhone applications for the assessment of active shoulder external rotation. Physiother Theory Pract 2014;30:521-5. http://dx.doi.org/10.3109/09593985.2014.900593
22. Mourcou Q, Fleury A, Diot B, Franco C, Vuillerme N. Mobile phone-based joint angle measurement for functional assessment and rehabilitation of proprioception. Biomed Res Int 2015;2015:328142. http://dx.doi.org/10.1155/2015/328142
23. Pichonnaz C, Duc C, Gleeson N, Ancey C, Jaccard H, Lécureux E, et al. Measurement properties of the smartphone-based B-B Score in current shoulder pathologies. Sensors (Basel) 2015;15:26801-17. http://dx.doi.org/10.3390/s151026801
24. Pohl M, Mehrholz J. A new shoulder range of motion screening measurement: its reliability and application in the assessment of the prevalence of shoulder contractures in patients with impaired consciousness caused by severe brain damage. Arch Phys Med Rehabil 2005;86:98-104. http://dx.doi.org/10.1016/j.apmr.2004.01.038
25. Russell TG, Jull GA, Wootton R. Can the Internet be used as a medium to evaluate knee angle? Man Ther 2003;8:242-6. http://dx.doi.org/10.1016/S1356-689X(03)00016-X
26. Shin SH, Ro du H, Lee OS, Oh JH, Kim SH. Within-day reliability of shoulder range of motion measurement with a smartphone. Man Ther 2012;17:298-304. http://dx.doi.org/10.1016/j.math.2012.02.010
27. Tekerekoğlu MS, Duman Y, Serindağ A, Cuğlan SS, Kaysadu H, Tunc E, et al. Do mobile phones of patients, companions and visitors carry multidrug-resistant hospital pathogens? Am J Infect Control 2011;39:379-81. http://dx.doi.org/10.1016/j.ajic.2010.10.026
28. Terwee CB, de Winter AF, Scholten RJ, Jans MP, Deville W, van Schaardenburg D, et al. Interobserver reproducibility of the visual estimation of range of motion of the shoulder. Arch Phys Med Rehabil 2005;86:1356-61. http://dx.doi.org/10.1016/j.apmr.2004.12.031
29. Werner BC, Holzgrefe RE, Griffin JW, Lyons ML, Cosgrove CT, Hart JM, et al. Validation of an innovative method of shoulder range-of-motion measurement using a smartphone clinometer application. J Shoulder Elbow Surg 2014;23:e275-82. http://dx.doi.org/10.1016/j.jse.2014.02.030
30. Williams CM, Caserta AJ, Haines TP. The TiltMeter app is a novel and accurate measurement tool for the weight bearing lunge test. J Sci Med Sport 2013;16:392-5. http://dx.doi.org/10.1016/j.jsams.2013.02.001

APPENDIX 2

HOOS AND WOMAC VALUES IN ASYMPTOMATIC AND ARTHRITIC COHORTS – IMPLICATIONS FOR LONGITUDINAL COHORT STUDIES

As published in Journal of Arthroplasty. In-press.

Hip Disability and Osteoarthritis Outcome Score and Western Ontario and McMaster Universities Osteoarthritis Index Values in Asymptomatic and Arthritic Cohorts

RAYMOND W. YU, MBBS, MSURGSCI; JAMES M. MCLEAN, MBBS, MS, FRACS; JASVIR S. BAHL, BAPPSC; LUCIAN B. SOLOMON, MD, PHD, FRACS

abstract

The primary aim of this study was to determine whether an electronic, multicenter data collection system could be used to establish normal population reference values for the Hip Disability and Osteoarthritis Outcome Score (HOOS) and the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC). The secondary aim was to investigate differences in asymptomatic HOOS and WOMAC values reported in 2 geographically distinct English-speaking countries and compare these with a symptomatic arthritic patient cohort. A total of 552 participants were recruited. Asymptomatic Australian and Canadian cohorts were compared; combined asymptomatic cohorts were compared with an arthritic cohort. There was a statistically significant association between age and asymptomatic HOOS ($P<.0001$) and WOMAC ($P<.0001$) values; as age increased, values worsened. Females had worse HOOS and WOMAC values ($P<.0001$). When compared with age- and sex-matched asymptomatic participants, arthritic participants had worse scores ($P<.0001$). Asymptomatic Australians had a statistically significant 3.8% better (higher) HOOS ($P<.0001$) in all age groups ($P<.0001$). When compared with age- and sex-matched asymptomatic participants, younger arthritic participants reported worse activities of daily living and sports and recreation HOOS values. This observational study established an electronic HOOS and WOMAC patient-reported outcome measures database of asymptomatic individuals in 2 geographically distinct countries. An asymptomatic control group should be sourced from the same country of origin as the proposed study. Factors that should be considered when recording the HOOS and WOMAC include age, sex, geographic location, history of an inactive hip problem, contralateral hip disease, and active knee, ankle, or foot problems. [*Orthopedics*. 201x; xx(x):xx-xx.]

The number of total hip arthroplasties (THAs) performed continues to increase.[1] Currently, most clinical research and joint registry outcome data involve observation of THA failure and revision rates. Patient-reported outcome measures (PROMs) allow for the evaluation of THA outcomes prior to

The authors are from the Orthopaedic and Trauma Service (RWY, JMM, LBS), Royal Adelaide Hospital, the Centre for Orthopaedic and Trauma Research (JMM, LBS), University of Adelaide, and the Alliance for Research in Exercise, Nutrition and Activity (JSB), Sansom Institute for Health Research & School of Health Sciences, University of South Australia, Adelaide, South Australia, Australia; and the Department of Orthopaedics (JMM), University of British Columbia, Vancouver, British Columbia, Canada.

The authors have no relevant financial relationships to disclose.

The authors thank Mami Okada, Research Assistant, University of British Columbia, for help establishing the study in Canada; Suzanne Edwards, Statistician, Data Management and Analysis Centre, University of Adelaide, for help with statistical analysis of the collected data; Tara-Louise McLean, Research Assistant to Dr James McLean, for help establishing the study in Australia and Canada; Catherine Hill, Professor of Rheumatology, University of Adelaide, for help establishing the study protocol; Tiffany Gill, Senior Research Fellow, University of Adelaide, for help establishing the study protocol; and Danny Awwad, Orthopaedic Registrar, Royal Adelaide Hospital, for help preparing the scattergrams used in this article.

Correspondence should be addressed to: James M. McLean, MBBS, MS, FRACS, Orthopaedic and Trauma Service, Royal Adelaide Hospital, 1 Port Rd, Adelaide, South Australia, Australia 5000 (james.mc_lean@adelaide.edu.au).

Received: July 31, 2018; Accepted: September 10, 2018.

doi: 10.3928/01477447-20190118-02

failure and revision. Patient-reported outcome measures have indicated that 15% of patients undergoing elective THA are not satisfied with their results.[2] These data are not routinely collected by most THA registries. Patient-reported outcome measures can provide surgeons and researchers with valuable long-term data on the severity of and change in symptoms before and after THA.

Approximately 20 different hip PROM scoring systems have been described.[3,4] The PROMs working group of the International Society of Arthroplasty Registries reported that the Hip Disability and Osteoarthritis Outcome Score (HOOS) and the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) are the PROMs most commonly used for THA registries.[5] The HOOS is a PROM used to measure subjective hip function and symptoms in adult patients with hip disability.[6,7] This includes individuals with or without hip osteoarthritis. The HOOS has the complete and original WOMAC within it.

The HOOS and the WOMAC can be administered quickly, requiring approximately 10 to 15 minutes for patients to complete. Swedish, Dutch, and French versions have been previously validated.[6,8] The Swedish and French versions of the HOOS have shown high responsiveness (ie, the ability to detect clinical change) before and after THA in patients with hip osteoarthritis.[4,7]

The HOOS and the WOMAC incorporate subjective reports of patients' hip and associated problems. Several investigators have reported an inverse relationship between age and PROM scores in asymptomatic populations, suggesting that as age increases, the reported "normal" PROM score decreases.[9] It is possible that a similar inverse relationship exists for the HOOS and the WOMAC; however, to the authors' knowledge, this has not been reported.

Despite their widespread use since their description in 1988 and 2003, no study has investigated the effect of administering an English version of the HOOS or the WOMAC to asymptomatic individuals and to individuals with symptomatic end-stage hip osteoarthritis planned for THA.[6,10] The primary aim of this study was to determine whether an electronic, multicenter data collection system could be used to establish normal, asymptomatic, population reference values for the HOOS and the WOMAC and to investigate potential regional differences in the reported values by assessing 2 geographically distinct English-speaking countries (Australia and Canada). The secondary aim was to compare this asymptomatic HOOS and WOMAC cohort with a symptomatic patient cohort with end-stage hip osteoarthritis planned for THA. The authors hypothesized that longitudinal data from symptomatic arthritic populations should be compared with data from age- and sex-matched control cohorts sourced from the same country of origin as the study (ie, local cohorts should only be compared with a local control group).

Materials and Methods

Independent ethics board approval was granted from each institution involved in this study. Adult participants were recruited from November 2014 to January 2017. Participants for the asymptomatic cohort were recruited from medical outpatient departments, driver's licensing offices, and community centers. There were no incentives or payment for participation. Informed consent was obtained from all individuals included in the study.

An asymptomatic cohort was established by approaching adult Australian and Canadian participants who identified themselves as having no symptoms of problems in either hip. Inclusion criteria were no self-reported pain or pathology in either hip and fluent in English. No medical notes or radiographs were reviewed for these individuals. Exclusion criteria were cognitive impairment, a history of inflammatory or hip arthritis, significant lumbar spine problems that interfered with function, active hip pathology, previous hip arthroplasty, or any hip surgery within the past 3 years. A history of inactive (ie, asymptomatic) hip pathology, including previous surgery more than 3 years earlier, was recorded. Reports of concurrent knee or ankle pathology or pain affecting function were recorded.

A consecutive sample of English-speaking patients with end-stage, radiographically confirmed osteoarthritis planned for THA were recruited from a single tertiary referral center. Exclusion criteria were cognitive impairment and significant lumbar spine problems interfering with function. A history of active knee or ankle pathology or pain affecting function, as well as a history of contralateral THA or hip problems, was recorded.

A total of 552 participants were recruited (496 asymptomatic and 56 symptomatic arthritic), with 273 being women and 279 being men. They were divided into 5 age subgroups: 40-49, 50-59, 60-69, 70-79, and 80 years and older.

Participants self-administered the HOOS (English version LK 2.0), which consisted of 40 items assessing 5 categories of PROMs: symptoms and stiffness (5 questions); pain (10 questions); function and daily living (17 questions); function, sports, and recreational activities (4 questions); and quality of life (4 questions). Participants were directed to answer each item by selecting 1 of 5 descriptive responses presented as a Likert scale. Each question was scored from 0 to 4, yielding a total score ranging from 0 (severe disability) to 100 (no disability).

The HOOS contained all WOMAC LK 3.0 questions in their original, unchanged form. The WOMAC scores were calculated as described by Nilsdotter et al[8] and included subscales for pain, stiffness, and function. All items were scored from 0 to 4, and scores for each of the subscales were calculated from the sum of the included items. A normalized score up to 100 was calculated for each subscale, with

0 indicating maximal problems and symptoms and 100 indicating no problems. A total WOMAC score of 0 to 100 was calculated for each participant, with a higher score indicating worse pain, stiffness, and function.

Chi-square and Fisher's exact tests were used for the analysis of differences between arthritic and asymptomatic cohorts. Ordinal logistic models were used for the analysis of differences in HOOS and WOMAC values because these scores are ordinal (ie, lower scores correlate with worse symptoms for HOOS, and higher scores correlate with worse symptoms for WOMAC). Poisson regression models were used to assess for associations between 5 HOOS and 4 WOMAC subscales and predictors such as country, age, and sex. SAS version 9.4 statistical software (SAS Institute Inc, Cary, North Carolina) was used.

Results

The demographics of each cohort are presented in **Table 1**. The HOOS values and statistical analyses are presented in **Table 2** and **Figure 1**. On adjustment for sex, asymptomatic Canadian participants had worse (lower) HOOS values across all age groups ($P<.0001$) compared with asymptomatic Australian participants. Arthritic participants had worse (lower) HOOS values than age- and sex-matched asymptomatic participants ($P<.0001$). Female participants had worse (lower) HOOS values than age-matched participants ($P<.0001$).

The HOOS subscale values and univariate Poisson regressions are presented in **Table 3**. Asymptomatic Australian participants had a HOOS for pain that was 2.1% better (higher) than that of age- and sex-matched asymptomatic Canadian participants ($P=.0123$; 95% confidence interval, 1.0044-1.0369). Asymptomatic Australian participants also had better (higher) HOOS subscale values for symptoms ($P=.0003$), activities of daily living ($P=.0006$), sports and recreation ($P<.0001$), and quality of life ($P<.0001$). Overall, asymptomatic Australian participants had a total HOOS value that was 3.8% better (higher) than that of age- and sex-matched asymptomatic Canadian participants ($P<.0001$).

The WOMAC scores and statistical analyses are presented in **Table 4** and **Figure 2**. On adjustment for sex, when compared with asymptomatic Australian participants, asymptomatic Canadian participants had worse (higher) WOMAC scores in the 70 to 79 years ($P=.0025$) and 80 years and older ($P<.0001$) age groups only. No difference in HOOS value was reported for asymptomatic participants younger than 70 years ($P>.05$). Arthritic participants had worse (higher) WOMAC scores than age- and sex-matched asymptomatic participants ($P<.0001$). Female participants had worse (higher) WOMAC scores compared with age-matched participants ($P<.0001$).

The WOMAC subscale scores and univariate Poisson regressions are presented in **Table 5**. Asymptomatic Australian participants had a better (lower) WOMAC score on all subscales compared with age- and sex-matched asymptomatic Canadian participants ($P<.0001$). Overall, asymptomatic Australian participants had a total WOMAC score that was 2.5% lower than that of age- and sex-matched asymptomatic Canadian participants ($P<.0001$).

Discussion

The HOOS and the WOMAC are PROMs that provide valuable information for surgeons and researchers assessing and quantifying outcomes before and after THA. This study assessed the HOOS and the WOMAC PROMs in asymptomatic and pathological, arthritic cohorts. Comparing both asymptomatic and arthritic hip cohorts enables a more accurate interpretation of the normal, age-related functional changes that will affect THA patients being followed on a longitudinal basis.

Establishment of a PROM reference database for asymptomatic individuals will aid in the interpretation of THA patients' PROM reports on a longitudinal basis. An electronic database allows remote administration, thereby minimizing resources, cost, and personnel required to collect, collate, and process data.[11] This theoretical advantage is enabled by the HOOS and the WOMAC PROMs containing only subjective responses and not requiring a face-to-face interaction with an assessor. These PROMs have the potential for remote access by both researchers and patients, who can complete questionnaires when convenient and have electronic reminders set. An electronic database also permits simultaneous data entry and data analyses, allows automated participant reminders or response time points, and has the potential to link to an electronic medical record. This could increase the response rate, which can be low with mailed paper questionnaires.

In this study, females in both the asymptomatic and the arthritic cohorts had worse HOOS and WOMAC values. This finding is similar to that of other studies.[12,13]

An inverse correlation was found between age and reported PROMs in asymptomatic participants (ie, as age increased, the reported PROM score was worse). This finding was not surprising given the age-related changes that occur with time and the accumulated medical and surgical comorbidities that can affect lower limb function and the resultant reported PROM score. Certainly, as patient age increases, perfect PROMs should not be expected.

Asymptomatic individuals 70 years and older showed great variation in reported HOOS and WOMAC values, with Canadian participants having worse PROM values than with their Australian counterparts. This trend was well demonstrated on comparison of each PROM graph, with national asymptomatic PROMs becoming more divergent in individuals 70 years and older. In contrast to the HOOS, which showed statistical differences between all age groups for all subscales, the WOMAC did not show statistically significant dif-

Table 1

Demographics of Australian Asymptomatic, Canadian Asymptomatic, and Arthritic Cohorts

Characteristic	Australian Asymptomatic Cohort (n=247)	Canadian Asymptomatic Cohort (n=249)	Arthritic Cohort (n=56)	Total (N=552)	Comparing Australian and Canadian Asymptomatic Cohorts	Comparing Asymptomatic and Arthritic Cohorts	Comparing All Cohorts
Sex, No.							
Male	123 (49.8%)	123 (49.4%)	33 (58.9%)	279 (50.5%)	*P*=.9290 (chi-square)	*P*=.4146 (chi-square)	*P*=.4146 (chi-square)
Female	124 (50.2%)	126 (50.6%)	23 (41.1%)	273 (49.5%)			
Age group, No.							
40-49 y	51 (20.7%)	53 (21.3%)	5 (8.9%)	109 (19.7%)			
50-59 y	72 (29.1%)	74 (29.7%)	8 (14.3%)	154 (27.9%)			
60-69 y	71 (28.7%)	70 (28.1%)	19 (33.9%)	160 (29.0%)	*P*=.9086 (chi-square)	*P*=.0128 (chi-square)	*P*=.0128 (chi-square)
70-79 y	33 (13.4%)	37 (14.9%)	16 (28.6%)	86 (15.6%)			
80+ y	20 (8.1%)	15 (6.0%)	8 (14.3%)	43 (7.8%)			
Age, mean (range), y	60.1 (40-90)	59.8 (40-94)	66.9 (43-89)	60.7 (40-94)			
History of an inactive (previous) hip problem (hip for asymptomatic cohort; contralateral hip for arthritic cohort), No.	0 (0%)	8 (3.2%)	24 (42.9%)	32 (5.8%)	*P*=.0074 (Fisher's exact test) (ie, asymptomatic Canadian cohort had a statistically significantly higher incidence of a history of an inactive [previous] hip problem compared with the asymptomatic Australian cohort)	*P*<.0001 (Fisher's exact test) (ie, arthritic cohort had a statistically significantly higher incidence of a contralateral history of an inactive [previous] hip problem compared with all asymptomatic participants)	*P*<.0001 (Fisher's exact test) (ie, arthritic cohort had a statistically significantly higher incidence of a contralateral history of an inactive [previous] hip problem compared with both asymptomatic cohorts assessed independently)
History of an active knee, ankle, and foot problem(s), No.	3 (1.2%)	39 (15.7%)	16 (28.6%)	58 (10.5%)	*P*<.0001 (Fisher's exact test) (ie, asymptomatic Canadian cohort had a statistically significantly higher incidence of an active knee, ankle, or foot problem[s] compared with the asymptomatic Australian cohort)	*P*<.0001 (Fisher's exact test) (ie, arthritic cohort had a statistically significantly higher incidence of an active knee, ankle, or foot problem[s] compared with all asymptomatic participants)	*P*<.0001 (Fisher's exact test) (ie, arthritic cohort had a statistically significantly higher incidence of an active knee, ankle, or foot problem[s] compared with both asymptomatic cohorts assessed independently)
Previous hip arthroplasty in contralateral hip, No.	NA as exclusion criteria	NA as exclusion criteria	20 (35.7%)	20 (3.6%)[a]		NA	

Abbreviation: NA, not applicable.
[a]*Included in arthritic cohort only by definition.*

Table 2

Hip Disability and Osteoarthritis Outcome Scores[a]

Characteristic	Asymptomatic Cohort, Mean HOOS			Arthritic Cohort Mean HOOS	Total Mean HOOS (Combined Asymptomatic and Arthritic Cohorts)	*P*		
	Australian	Canadian	Combined			Comparing Australian and Canadian Asymptomatic Cohorts	Comparing Combined Asymptomatic and Arthritic Cohorts	Comparing All Cohorts
Sex								
Male	97.28	92.52	94.90	29.68	87.19	.0002	<.0001	<.0001
Female	95.95	90.86	93.39	23.84	87.53	<.0001	<.0001	<.0001
Age group								
40-49 y	98.78	94.98	96.84	4.88	93.08	.0383	<.0001	<.0001
50-59 y	97.88	94.47	96.15	19.63	92.18	.0271	<.0001	<.0001
60-69 y	94.63	91.09	92.88	24.45	84.75	.0243	<.0001	<.0001
70-79 y	94.69	88.25	91.29	34.96	80.81	.0039	<.0001	<.0001
80+ y	96.77	77.45	88.49	34.08	78.37	<.0001	<.0001	<.0001
Overall	96.62	91.68	94.14	27.29	87.36	<.0001	<.0001	.0001

Abbreviation: HOOS, Hip Disability and Osteoarthritis Outcome Score.
[a]Maximum score of 100 points, with a lower score indicating worse pain, stiffness, and functional limitation.

ferences for all subscales. The WOMAC did show a significant difference for individuals 70 years and older, but not for younger individuals. This observation may be explained by an increased relevance of these particular subscale questions to younger and more high-demand individuals, as the HOOS was developed as an expansion of the WOMAC (with the inclusion of questions relating to sports and recreation and decreased quality of life). Therefore, the authors recommend that the HOOS be used in preference to the WOMAC to detect differences between cohorts in the younger age groups (40 to 69 years).

Arthritic participants reported statistically worse PROMs compared with their age- and sex-matched asymptomatic counterparts. This finding was not surprising given the end-stage arthritic changes that defined this cohort and the effect that this had on their lower limb function. Interestingly, in contrast to the asymptomatic cohort, in this study a correlation was found between age and reported PROM in the arthritic cohort (ie, as age increased, PROM score improved). The authors postulate that the apparent age-related improvement in PROMs was unlikely to be related to the

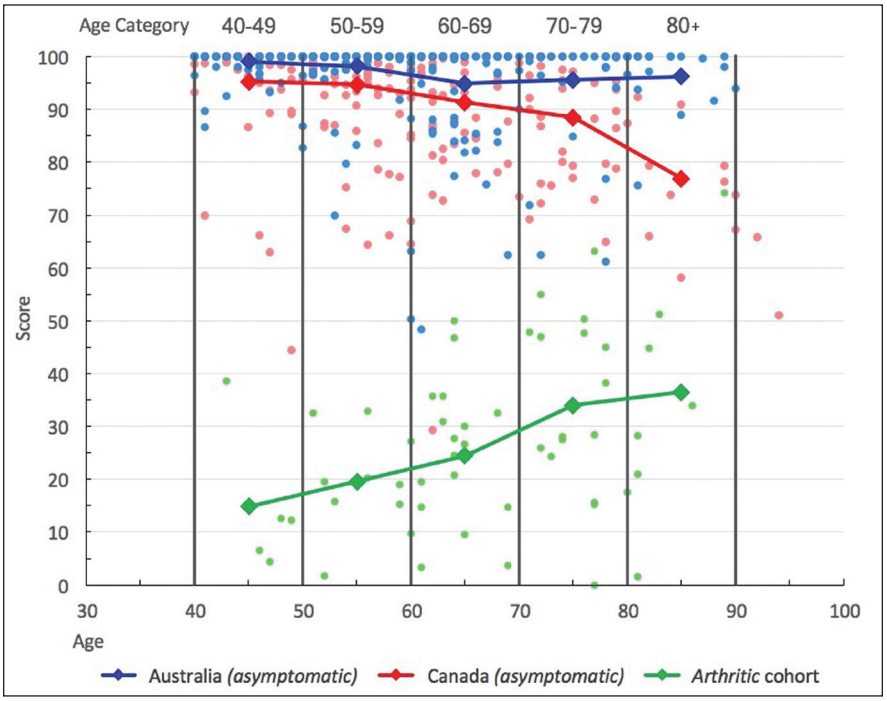

Figure 1: Scattergram showing Hip Disability and Osteoarthritis Outcome Score for each age group among asymptomatic Australian participants (blue), asymptomatic Canadian participants (red), and arthritic cohort (green). Higher scores indicate better outcomes for symptoms, pain, function, and quality of life.

Table 3

Univariate Poisson Regressions of Hip Disability and Osteoarthritis Outcome Score Subscale Scores

	Asymptomatic Cohort, Mean Score			
HOOS Subscale	Australian	Canadian	95% Confidence Interval	*P*
Pain	97.67	95.71	1.0044-1.0369	.0123
Symptoms	97.23	94.44	1.0133-1.0462	.0003
Activities of daily living	97.93	95.24	1.0120-1.0447	.0006
Sports and recreation	94.74	87.40	1.0664-1.1019	<.0001
Quality of life	96.25	92.63	1.0225-1.0559	<.0001
Total score	483.83	465.41	1.0321-1.0470	<.0001

Abbreviation: HOOS, Hip Disability and Osteoarthritis Outcome Score.

natural history of osteoarthritis, but rather to the threshold of operative intervention for individuals with radiographically confirmed osteoarthritis. Because, by definition, all of the arthritic participants were indicated and planned for surgery, younger individuals in this cohort required worse symptoms and greater impact on function before reaching the threshold for consideration for THA intervention. Older arthritic individuals likely had a lower threshold indication for THA, given the lower implant demands and lower risk of revision surgery in these age groups.[1]

Younger arthritic participants were also more likely to report worse activities of daily living and sports and recreation HOOS values than their asymptomatic age-matched counterparts. This study also found that as age increased, the disparity between scores in all cohorts diminished, particularly the subscales (ie, as age increases, asymptomatic and symptomatic PROM values approach one another). This observation highlights the implications of osteoarthritis in younger arthritic individuals, which likely translate to a greater impact on occupation, livelihood, and recreation. Future longitudinal studies should assess the impact of THA on reported PROMs in younger individuals compared with the more common elderly THA patients.

Factors other than age were also found to be associated with differences in PROM values. Participants who reported a history of an inactive (previous) hip problem or an active knee, ankle, or foot problem had worse PROM scores. The asymptomatic Canadian cohort reported a higher inci-

Table 4

Western Ontario and McMaster Universities Osteoarthritis Index Scores[a]

	Asymptomatic Cohort, Mean Score					*P*		
Characteristic	Australian	Canadian	Combined	Arthritic Cohort Mean Score	Total Mean Score	Comparing Australia and Canadian Asymptomatic Cohorts	Comparing Combined Asymptomatic and Arthritic Cohorts	Comparing All Cohorts
Sex								
Male	1.79	4.97	3.36	61.79	10.32	.0018	<.0001	<.0001
Female	2.80	6.38	4.60	68.43	9.98	.0004	<.0001	<.0001
Age group								
40-49 y	0.96	3.25	2.13	80.60	5.72	.1237	<.0001	<.0001
50-59 y	1.47	3.42	2.44	70.25	6.01	.1231	<.0001	<.0001
60-69 y	3.65	5.86	4.74	65.32	11.94	.0830	<.0001	<.0001
70-79 y	3.36	8.84	6.26	57.63	15.81	.0025	<.0001	<.0001
80+ y	2.10	16.67	8.34	60.63	18.07	<.0001	<.0001	<.0001
Overall	2.30	5.69	3.99	64.52	10.15	<.0001	<.0001	.0001

[a]A total of 24 items with a possible maximum score of 96. The final score is expressed in a percentage and calculated by dividing an individual's score by the total score and multiplying that by 100, thereby making a normalized maximum 100-point score. A higher score indicates worse pain, stiffness, and functional limitation.

dence of inactive hip problems, as well as more active knee, ankle, or foot problems. This may also have influenced the PROM values reported, as both of these variables were associated with worse PROM scores in the combined asymptomatic cohorts, independent of age.

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 for patients with multiple, concurrent lower limb pathologies. To the authors' knowledge, few (if any) studies investigating either asymptomatic or arthritic cohorts have considered the effect these variables may have on the reported HOOS and WOMAC values. In addition, most studies do not record the incidence of contralateral hip disease (either active or inactive), concurrent spinal disease, or active bilateral or unilateral knee, ankle, or foot pathology. Rarely is a history of contralateral hip disease or THA reported in arthritic or THA patient cohorts.[8,14-16]

In this study, efforts were made beyond those reported previously to establish an asymptomatic control group that fulfilled the criteria of being a control group. Exclusion criteria were carefully chosen to minimize the inclusion of participants with active pathology and to collect an asymptomatic population who identified themselves as having normal hips. To the authors' knowledge, this study represents the largest database of normal HOOS and WOMAC values reported in the literature. Although other researchers have recorded normal values for other musculoskeletal assessment tools, most have collected fewer than 150 participants[17] or have studied only young, active individuals[18] or individuals not first screened for active joint pathology.[13]

Several studies have indicated that a perfect hip score may not be reported in an asymptomatic or disease-free population.[13,19] In a representative sample, it is assumed that a certain percentage of the general population will have asymptomatic, incidental hip pathology. A representative control cohort without active pathology would include a small percentage of participants with the following: a history of inactive hip pathology; a history of previous successful non-arthroplasty hip surgery; incidental subclinical hip pathology; and incidental clinical hip pathology (undiagnosed or not investigated). In this study, all efforts within the constraints of ethics approval, budget, and practical participant selection were made to exclude patients with active hip pathology from the asymptomatic cohort. Many previous

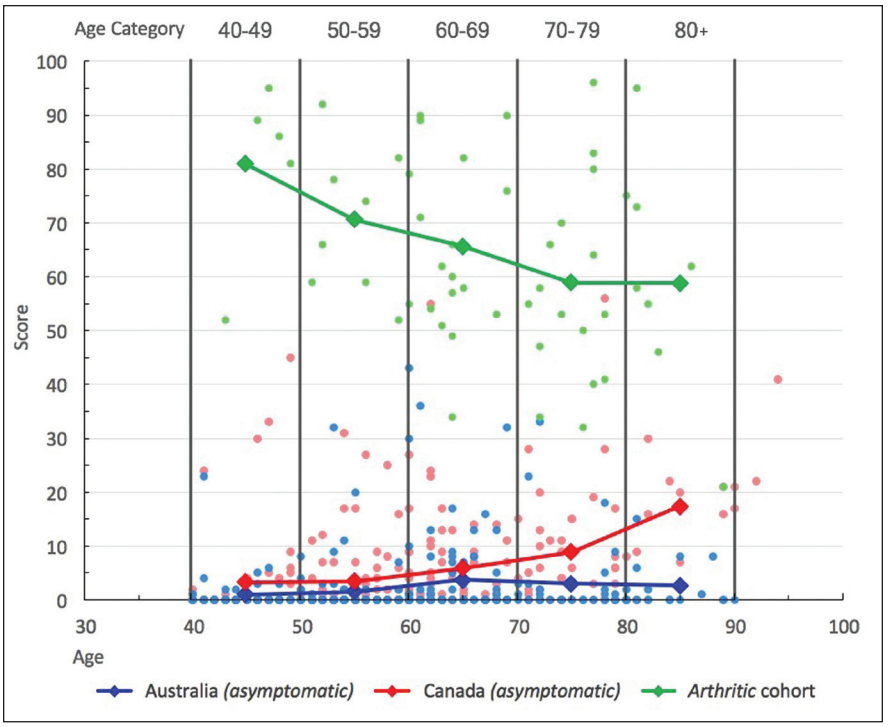

Figure 2: Scattergram showing Western Ontario and McMaster Universities Osteoarthritis Index score for each age group among asymptomatic Australian participants (blue), asymptomatic Canadian participants (red), and arthritic cohort (green). Higher scores indicate worse pain, stiffness, and function.

Table 5

Univariate Poisson Regressions of Western Ontario and McMaster Universities Osteoarthritis Index Subscale Scores

WOMAC Subscale	Asymptomatic Cohort, Mean Score: Australian	Asymptomatic Cohort, Mean Score: Canadian	95% Confidence Interval	*P*
Pain	0.39	0.76	0.42-0.52	<.0001
Stiffness	0.24	0.53	0.35-0.45	<.0001
Function	1.43	3.34	0.38-0.43	<.0001
Total score	2.07	4.63	0.41-0.45	<.0001

Abbreviation: WOMAC, Western Ontario and McMaster Universities Osteoarthritis Index.

studies investigating normal reference values did not attempt to exclude these participants from their sample cohorts.[13] Thus, in these studies, the numbers of responders with active hip pathology or a history of hip pathology are unknown. The authors recommend that future studies consider these variables when establishing a control group against which to compare an interventional group using the HOOS and WOMAC PROMs.

This study reported differences between 2 geographically distinct asymptomatic cohorts from 2 English-speaking Western countries. Although this study represents the largest cohort of asymptomatic hip PROM scores reported in the literature, this finding may reflect a true difference between the population cohorts or a selection or reporting bias. Cultural and societal differences may also explain the disparity in scores. Although the distribution of ethnicity was not explored in this study, the 2 countries have Westernized societies and similar distributions based on reports from the Australian Bureau of Statistics[20] and Statistics Canada.[21] Future studies with larger samples may be able to assess these reported differences. On the basis of their findings, the authors recommend that future studies using the HOOS and the WOMAC be performed with comparisons made against a control group sourced from the same country of origin.

The current study had important limitations that should be considered when interpreting the results. The arthritic cohort represented a sample of consecutive patients with end-stage osteoarthritis planned for THA. In contrast, the asymptomatic cohort was collected randomly, with no specific randomization method regarding participant identification being employed. Because no radiographs were obtained to confirm whether participants had asymptomatic degenerative hip disease, it is possible that some of these individuals were included in the cohorts even though they self-reported as asymptomatic.

As with any observational study, there is the potential for selection bias, particularly when there is no randomization. The primary benefit of randomization is the elimination of both conscious and unconscious bias associated with the selection of a participant. Randomization was not employed in this study secondary to time, resource, and cost constraints.

Another potential source of selection bias involves the use of electronic questionnaires. Patients with severe wrist and hand pathology may be less able to complete electronic questionnaires using computers or tablets. The technology also assumes that the patient is capable of being and willing to be an active participant and that the patient prefers a virtual method of follow-up over a face-to-face clinician–patient interaction. This may not be the case for the entire population and especially for the elderly, who may not be computer savvy.[22] Participants may have declined involvement because of the technology. Interviewer bias may have been introduced, especially when elderly participants were assisted with using the technology.

Asymptomatic 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 the authors chose to exclude participants with active hip disease, they did include participants with a history of a previous hip problem that "no longer bothered them" (ie, asymptomatic). Because this is a purely subjective report, it is possible that some asymptomatic individuals with a prior hip problem and only minor functional incapacities who could have been included self-excluded themselves.

The arthritic cohort was sourced from Australia, and from the same city as the asymptomatic Australian cohort. Although differences in PROM values in asymptomatic cohorts from 2 geographically distinct English-speaking Western countries were reported, no such comparisons could be made for the arthritic cohort. Therefore, future THA studies should involve comparison with a control group sourced from the same country of origin.

Because the main goal of this study was the establishment of an asymptomatic control group, longitudinal data were not collected over several time points; hence, response rate could not be assessed and the PROMs could not be validated. Validation, which requires collection of multiple data over different time points, was not within the scope of this study. Formal validity studies have not been published for English translations of the HOOS and the WOMAC. Future studies could collect the WOMAC and the HOOS values in both asymptomatic and arthritic cohorts on a longitudinal basis. However, this would require significant administration, resources, and personnel. A positive step toward this goal is the establishment of an electronic database of population reference values for various ages across the sexes for the HOOS and the WOMAC PROMs.

To the authors' knowledge, this study represents the largest cohort of combined asymptomatic and arthritic participants using the HOOS and the WOMAC PROMs. The authors aim to expand this database to include an arthritic population of equivalent size with longitudinal data. The HOOS and the WOMAC are preferable in the development of a large electronic catalog of normal control population data because they permit remote administration. With resource management and cost justification becoming more of a focus in health care, PROMs that can be administered remotely make future research, especially international multicenter cohort studies, a possibility.

Conclusion

This observational study established an electronic HOOS and WOMAC PROM database of asymptomatic individuals in 2 geographically distinct English-speaking countries for comparison with age- and sex-matched arthritic and THA patients. The HOOS should be used in preference to the WOMAC when assessing for

variation in younger age groups (40 to 69 years). An asymptomatic control group should ideally be sourced from the same country of origin as the proposed study. Factors that should be considered when recording the HOOS and the WOMAC include age, sex, geographic location, a history of an inactive (previous) hip problem, contralateral hip disease, and an active knee, ankle, or foot problem.

References

1. Australian Orthopaedic Association National Joint Replacement Registry. *Hip, Knee & Shoulder Arthroplasty: Annual Report 2017.* Adelaide, South Australia, Australia: Australian Orthopaedic Association; 2017.
2. Rothwell A, Oakley A, Larmer P, Hobbs T. *The New Zealand Joint Registry Seventeen Year Report.* Wellington, New Zealand: Orthopaedic Association; 2016.
3. Bryant MJ, Kernohan WG, Nixon JR, Mollan RA. A statistical analysis of hip scores. *J Bone Joint Surg Br*. 1993;75(5):705-709.
4. 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. *Arthritis Care Res (Hoboken).* 2011;63(suppl 11):S200-S207.
5. Rolfson O, Eresian Chenok K, Bohm E, et al; and Patient-Reported Outcome Measures Working Group of the International Society of Arthroplasty Registries. Patient-reported outcome measures in arthroplasty registries. *Acta Orthop*. 2016;87(suppl 1):3-8.
6. 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.
7. 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.
8. 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;4:10.
9. McLean JM, Cappelletto J, Clarnette J, et al. Normal population reference values for the Oxford and Harris Hip Scores: electronic data collection and its implications for clinical practice. *Hip Int*. 2017;27(4):389-396.
10. 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.
11. Bellamy N, Wilson C, Hendrikz J, et al. Osteoarthritis Index delivered by mobile phone (m-WOMAC) is valid, reliable, and responsive. *J Clin Epidemiol*. 2011;64(2):182-190.
12. Mannion AF, Impellizzeri FM, Naal FD, Leunig M. Women demonstrate more pain and worse function before THA but comparable results 12 months after surgery. *Clin Orthop Relat Res*. 2015;473(12):3849-3857.
13. Bellamy N, Wilson C, Hendrikz J. Population-based normative values for the Western Ontario and McMaster (WOMAC) Osteoarthritis Index: part I. *Semin Arthritis Rheum*. 2011;41(2):139-148.
14. Mehta SP, Sankar A, Venkataramanan V, et al. Cross-cultural validation of the ICOAP and physical function short forms of the HOOS and KOOS in a multi-country study of patients with hip and knee osteoarthritis. *Osteoarthritis Cartilage*. 2016;24(12):2077-2081.
15. Alzahrani MM, Smith K, Tanzer D, Tanzer M. Primary total hip arthroplasty: equivalent outcomes in low and high functioning patients. *J Am Acad Orthop Surg*. 2016;24(11):814-822.
16. Davis AM, Perruccio AV, Canizares M, et al. Comparative, validity and responsiveness of the HOOS-PS and KOOS-PS to the WOMAC physical function subscale in total joint replacement for osteoarthritis. *Osteoarthritis Cartilage*. 2009;17(7):843-847.
17. 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.
18. 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.
19. Saleh KJ, Davis A. Measures for pain and function assessments for patients with osteoarthritis. *J Am Acad Orthop Surg*. 2016;24(11):e148-e162.
20. Australian Bureau of Statistics. Annual report 2016-17. http://www.ausstats.abs.gov.au/Ausstats/subscriber.nsf/0/3B641F80D47793B2CA2581BD000F8996/$File/10010_2016_17.pdf. Accessed February 4, 2018.
21. Canada CGo. Ethnic origins for Canada, provinces and territories Census of Population. 2011. https://www12.statcan.gc.ca/nhs-enm/2011/ref/guides/99-010-x/99-010-x2011006-eng.pdf. Accessed July 10, 2017.
22. Snyder CF, Aaronson NK, Choucair AK, et al. Implementing patient-reported outcomes assessment in clinical practice: a review of the options and considerations. *Qual Life Res*. 2012;21(8):1305-1314.

Bibliography

1. Ehrenpreis ED, Zhou Y. Hospital costs, length of stay and prevalence of hip and knee arthroplasty in patients with inflammatory bowel disease. World J Gastroenterol. 2017 Jul 14;23(26):4752-8. DOI: 10.3748/wjg.v23.i26.4752.
2. Poitras S, Au K, Wood K, Dervin G, Beaule PE. Predicting hospital length of stay and short-term function after hip or knee arthroplasty: are both performance and comorbidity measures useful? Int Orthop. 2018 Oct;42(10):2295-300. DOI: 10.1007/s00264-018-3833-y.
3. Tilbury C, Haanstra TM, Verdegaal SHM, Nelissen R, de Vet HCW, Vliet Vlieland TPM, et al. Patients' pre-operative general and specific outcome expectations predict postoperative pain and function after total knee and total hip arthroplasties. Scand J Pain. 2018 Jul 26;18(3):457-66. DOI: 10.1515/sjpain-2018-0022.
4. Levine DM, Morlock LL, Mushlin AI, Shapiro S, Malitz FE. The role of new health practitioners in a prepaid group practice: provider differences in process and outcomes of medical care. Med Care. 1976 Apr;14(4):326-47.
5. Vincent HK, Hagen JE, Zdziarski-Horodyski LA, Patrick M, Sadasivan KK, Guenther R, et al. Patient-Reported Outcomes Measurement Information System Outcome Measures and Mental Health in Orthopaedic Trauma Patients During Early Recovery. J Orthop Trauma. 2018 Sep;32(9):467-73. DOI: 10.1097/BOT.0000000000001245.
6. Siljander MP, McQuivey KS, Fahs AM, Galasso LA, Serdahely KJ, Karadsheh MS. Current Trends in Patient-Reported Outcome Measures in Total Joint Arthroplasty: A Study of 4 Major Orthopaedic Journals. J Arthroplasty. 2018 Nov;33(11):3416-21. DOI: 10.1016/j.arth.2018.06.034.
7. Wongworawat MD. Editor's Spotlight/Take 5: Readability of Orthopaedic Patient-reported Outcome Measures: Is There a Fundamental Failure to Communicate? Clin Orthop Relat Res. 2017 Aug;475(8):1932-5. DOI: 10.1007/s11999-017-5383-9.
8. Perez JL, Mosher ZA, Watson SL, Sheppard ED, Brabston EW, McGwin G, Jr., et al. Readability of Orthopaedic Patient-reported Outcome Measures: Is There a Fundamental Failure to Communicate? Clin Orthop Relat Res. 2017 Aug;475(8):1936-47. DOI: 10.1007/s11999-017-5339-0.
9. 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 Jul;80(4):600-6.
10. Dawson J, Fitzpatrick R, Murray D, Carr A. Questionnaire on the perceptions of patients about total knee replacement. J Bone Joint Surg Br. 1998 Jan;80(1):63-9.
11. MacDermid JC. Development of a scale for patient rating of wrist pain and disability. J Hand Ther. 1996 Apr-Jun;9(2):178-83.

12. Hudak PL, Amadio PC, Bombardier C. Development of an upper extremity outcome measure: the DASH (disabilities of the arm, shoulder and hand) [corrected]. The Upper Extremity Collaborative Group (UECG). Am J Ind Med. 1996 Jun;29(6):602-8. DOI: 10.1002/(SICI)1097-0274(199606)29:6<602::AID-AJIM4>3.0.CO;2-L.
13. Dawson J, Fitzpatrick R, Carr A, Murray D. Questionnaire on the perceptions of patients about total hip replacement. J Bone Joint Surg Br. 1996 Mar;78(2):185-90.
14. Dawson J, Fitzpatrick R, Carr A. Questionnaire on the perceptions of patients about shoulder surgery. J Bone Joint Surg Br. 1996 Jul;78(4):593-600.
15. Conboy VB, Morris RW, Kiss J, Carr AJ. An evaluation of the Constant-Murley shoulder assessment. J Bone Joint Surg Br. 1996 Mar;78(2):229-32.
16. Richards RR, An KN, Bigliani LU, Friedman RJ, Gartsman GM, Gristina AG, et al. A standardized method for the assessment of shoulder function. J Shoulder Elbow Surg. 1994 Nov;3(6):347-52. DOI: 10.1016/S1058-2746(09)80019-0.
17. Roach KE, Budiman-Mak E, Songsiridej N, Lertratanakul Y. Development of a shoulder pain and disability index. Arthritis care and research : the official journal of the Arthritis Health Professions Association. 1991 Dec;4(4):143-9.
18. Insall JN, Dorr LD, Scott RD, Scott WN. Rationale of the Knee Society clinical rating system. Clin Orthop Relat Res. 1989 Nov(248):13-4.
19. Ewald FC. The Knee Society total knee arthroplasty roentgenographic evaluation and scoring system. Clin Orthop Relat Res. 1989 Nov(248):9-12.
20. Ellman H, Hanker G, Bayer M. Repair of the rotator cuff. End-result study of factors influencing reconstruction. J Bone Joint Surg Am. 1986 Oct;68(8):1136-44.
21. Amstutz HC, Sew Hoy AL, Clarke IC. UCLA anatomic total shoulder arthroplasty. Clin Orthop Relat Res. 1981 Mar-Apr(155):7-20.
22. Ilstrup DM, Nolan DR, Beckenbaugh RD, Coventry MB. Factors influencing the results in 2,012 total hip arthroplasties. Clin Orthop Relat Res. 1973 Sep(95):250-62.
23. 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 Jun;51(4):737-55.
24. Coons SJ, Gwaltney CJ, Hays RD, Lundy JJ, Sloan JA, Revicki DA, et al. Recommendations on evidence needed to support measurement equivalence between electronic and paper-based patient-reported outcome (PRO) measures: ISPOR ePRO Good Research Practices Task Force report. Value Health. 2009 Jun;12(4):419-29. DOI: 10.1111/j.1524-4733.2008.00470.x.
25. Gomez-Valero S, Garcia-Perez F, Florez-Garcia MT, Miangolarra-Page JC. Assessment of cross-cultural adaptations of patient-reported shoulder outcome measures in Spanish: a systematic review. Shoulder Elbow. 2017 Oct;9(4):233-46. DOI: 10.1177/1758573217694340.

26. 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 Apr;83(3):339-44.
27. Lingard EA, Katz JN, Wright RJ, Wright EA, Sledge CB, Kinemax Outcomes G. Validity and responsiveness of the Knee Society Clinical Rating System in comparison with the SF-36 and WOMAC. J Bone Joint Surg Am. 2001 Dec;83-A(12):1856-64.
28. Murray DW, Fitzpatrick R, Rogers K, Pandit H, Beard DJ, Carr AJ, et al. The use of the Oxford hip and knee scores. J Bone Joint Surg Br. 2007 Aug;89(8):1010-4. DOI: 10.1302/0301-620X.89B8.19424.
29. 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 Aug;28(2):88-96. DOI: 10.2519/jospt.1998.28.2.88.
30. Roos EM, Roos HP, Ekdahl C, Lohmander LS. Knee injury and Osteoarthritis Outcome Score (KOOS)--validation of a Swedish version. Scand J Med Sci Sports. 1998 Dec;8(6):439-48.
31. Silva AL, Demange MK, Gobbi RG, da Silva TF, Pecora JR, Croci AT. Translation and Validation of the Knee Society Score - KSS for Brazilian Portuguese. Acta Ortop Bras. 2012;20(1):25-30. DOI: 10.1590/S1413-78522012000100005.
32. Harcourt WG, White SH, Jones P. Specificity of the Oxford knee status questionnaire. The effect of disease of the hip or lumbar spine on patients' perception of knee disability. J Bone Joint Surg Br. 2001 Apr;83(3):345-7.
33. 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 Aug;28(3):367-94. DOI: 10.1016/j.cger.2012.05.007.
34. Thorborg K, Tijssen M, Habets B, Bartels EM, Roos EM, Kemp J, 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 Jun;49(12):812. DOI: 10.1136/bjsports-2014-094224.
35. 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 Mar;48(6):458-63. DOI: 10.1136/bjsports-2012-092072.
36. Klassbo 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.
37. Ware J, Jr., Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care. 1996 Mar;34(3):220-33.

38. Kemp JL, Collins NJ, Roos EM, Crossley KM. Psychometric properties of patient-reported outcome measures for hip arthroscopic surgery. Am J Sports Med. 2013 Sep;41(9):2065-73. DOI: 10.1177/0363546513494173.
39. Soderman P, Malchau H, Herberts P. Outcome of total hip replacement: a comparison of different measurement methods. Clin Orthop Relat Res. 2001 Sep(390):163-72.
40. Aram P, Trela-Larsen L, Sayers A, Hills AF, Blom AW, McCloskey EV, et al. Estimating an Individual's Probability of Revision Surgery After Knee Replacement: A Comparison of Modeling Approaches Using a National Data Set. Am J Epidemiol. 2018 Oct 1;187(10):2252-62. DOI: 10.1093/aje/kwy121.
41. Haase E, Kopkow C, Beyer F, Lutzner J, Kirschner S, Hartmann A, et al. Patient-reported outcomes and outcome predictors after primary total hip arthroplasty: results from the Dresden Hip Surgery Registry. Hip Int. 2016 Jan-Feb;26(1):73-81. DOI: 10.5301/hipint.5000300.
42. Paulsen A. Patient reported outcomes in hip arthroplasty registries. Dan Med J. 2014 May;61(5):B4845.
43. Alviar MJ, Olver J, Brand C, Tropea J, Hale T, Pirpiris M, et al. Do patient-reported outcome measures in hip and knee arthroplasty rehabilitation have robust measurement attributes? A systematic review. Journal of rehabilitation medicine. 2011 Jun;43(7):572-83. DOI: 10.2340/16501977-0828.
44. Mygind-Klavsen B, Lund B, Nielsen TG, Maagaard N, Kraemer O, Holmich P, et al. Danish Hip Arthroscopy Registry: predictors of outcome in patients with femoroacetabular impingement (FAI). Knee Surg Sports Traumatol Arthrosc. 2018 Apr 25. DOI: 10.1007/s00167-018-4941-3.
45. Lund B, Mygind-Klavsen B, Gronbech Nielsen T, Maagaard N, Kraemer O, Holmich P, et al. Danish Hip Arthroscopy Registry (DHAR): the outcome of patients with femoroacetabular impingement (FAI). J Hip Preserv Surg. 2017 Jul;4(2):170-7. DOI: 10.1093/jhps/hnx009.
46. Pearse AJ, Hooper GJ, Rothwell A, Frampton C. Survival and functional outcome after revision of a unicompartmental to a total knee replacement: the New Zealand National Joint Registry. J Bone Joint Surg Br. 2010 Apr;92(4):508-12. DOI: 10.1302/0301-620X.92B4.22659.
47. Dawson J, Rogers K, Fitzpatrick R, Carr A. The Oxford shoulder score revisited. Arch Orthop Trauma Surg. 2009 Jan;129(1):119-23. DOI: 10.1007/s00402-007-0549-7.
48. Clarke MG, Dewing CB, Schroder DT, Solomon DJ, Provencher MT. Normal shoulder outcome score values in the young, active adult. J Shoulder Elbow Surg. 2009 May-Jun;18(3):424-8. DOI: 10.1016/j.jse.2008.10.009.
49. MacDermid JC, Solomon P, Prkachin K. The Shoulder Pain and Disability Index demonstrates factor, construct and longitudinal validity. BMC Musculoskelet Disord. 2006;7:12. DOI: 10.1186/1471-2474-7-12.
50. Dawson J, Fitzpatrick R, Carr A. The assessment of shoulder instability. The development and validation of a questionnaire. J Bone Joint Surg Br. 1999 May;81(3):420-6.

51. Lieberman JR, Hawker G, Wright JG. Hip function in patients >55 years old: population reference values. J Arthroplasty. 2001 Oct;16(7):901-4. DOI: 10.1054/arth.2001.26593.
52. Hayes K, Walton JR, Szomor ZR, Murrell GA. Reliability of five methods for assessing shoulder range of motion. The Australian journal of physiotherapy. 2001;47(4):289-94.
53. Constant CR, Gerber C, Emery RJ, Sojbjerg JO, Gohlke F, Boileau P. A review of the Constant score: modifications and guidelines for its use. J Shoulder Elbow Surg. 2008 Mar-Apr;17(2):355-61. DOI: 10.1016/j.jse.2007.06.022.
54. Clement ND, Court-Brown CM. Oxford shoulder score in a normal population. Int J Shoulder Surg. 2014 Jan;8(1):10-4. DOI: 10.4103/0973-6042.131849.
55. Katolik LI, Romeo AA, Cole BJ, Verma NN, Hayden JK, Bach BR. Normalization of the Constant score. J Shoulder Elbow Surg. 2005 May-Jun;14(3):279-85. DOI: 10.1016/j.jse.2004.10.009.
56. Noorani AM, Roberts DJ, Malone AA, Waters TS, Jaggi A, Lambert SM, et al. Validation of the Stanmore percentage of normal shoulder assessment. Int J Shoulder Surg. 2012 Jan;6(1):9-14. DOI: 10.4103/0973-6042.94307.
57. Plaschke HC, Jorgensen A, Thillemann TM, Brorson S, Olsen BS. Validation of the Danish version of the Oxford Elbow Score. Dan Med J. 2013 Oct;60(10):A4714.
58. Romero L, Nieuwenhuijse M, Carr A, Sedrakyan A. Review of clinical outcomes-based anchors of minimum clinically important differences in hip and knee registry-based reports and publications. J Bone Joint Surg Am. 2014 Dec 17;96 Suppl 1:98-103. DOI: 10.2106/JBJS.N.00464.
59. Griffiths-Jones W, Norton MR, Fern ED, Williams DH. The Equivalence of Remote Electronic and Paper Patient Reported Outcome (PRO) Collection. J Arthroplasty. 2014 Jul 4. DOI: 10.1016/j.arth.2014.07.003.
60. Rolfson O, Wissig S, van Maasakkers L, Stowell C, Ackerman I, Ayers D, et al. Defining an International Standard Set of Outcome Measures for Patients With Hip or Knee Osteoarthritis: Consensus of the International Consortium for Health Outcomes Measurement Hip and Knee Osteoarthritis Working Group. Arthritis Care Res (Hoboken). 2016 Nov;68(11):1631-9. DOI: 10.1002/acr.22868.
61. Registry AOANJR. Other Registries Worldwide. Adelaide, SA: Data Management & Analysis Centre, University of Adelaide; 2015. Available from: https://aoanjrr.dmac.adelaide.edu.au/registries
62. Roos EM, Toksvig-Larsen S. Knee injury and Osteoarthritis Outcome Score (KOOS) - validation and comparison to the WOMAC in total knee replacement. Health and quality of life outcomes. 2003;1:17. DOI: 10.1186/1477-7525-1-17.

63. Williamson T, Sikka R, Tompkins M, Nelson BJ. Use of the Knee Injury and Osteoarthritis Outcome Score in a Healthy United States Population. Am J Sports Med. 2016 Feb;44(2):440-6. DOI: 10.1177/0363546515616812.
64. Bremner-Smith AT, Ewings P, Weale AE. Knee scores in a 'normal' elderly population. Knee. 2004 Aug;11(4):279-82. DOI: 10.1016/j.knee.2003.06.001.
65. Raja A, Williamson T, Horst PK. Extrapolation of Normative KOOS, JR Data for the Young Patient Population Undergoing Knee Arthroplasty Procedures. J Arthroplasty. 2018 Dec;33(12):3655-9. DOI: 10.1016/j.arth.2018.08.031.
66. Hamilton DF, Giesinger JM, Patton JT, MacDonald DJ, Simpson AH, Howie CR, et al. Making the Oxford Hip and Knee Scores meaningful at the patient level through normative scoring and registry data. Bone Joint Res. 2015 Aug;4(8):137-44. DOI: 10.1302/2046-3758.48.2000524.
67. El-Boghdadly K, Abdallah FW, Short A, Vorobeichik L, Memtsoudis SG, Chan VWS. Outcome Selection and Methodological Quality of Major and Minor Shoulder Surgery Studies: A Scoping Review. Clin Orthop Relat Res. 2019 Mar;477(3):606-19. DOI: 10.1097/CORR.0000000000000578.
68. Haviv B, Rutenberg TF, Bronak S, Yassin M. Arthroscopic rotator cuff surgery following shoulder trauma improves outcome despite additional pathologies and slow recovery. Knee Surg Sports Traumatol Arthrosc. 2018 Dec;26(12):3804-9. DOI: 10.1007/s00167-018-4969-4.
69. Unger RZ, Burnham JM, Gammon L, Malempati CS, Jacobs CA, Makhni EC. The Responsiveness of Patient- Reported Outcome Tools in Shoulder Surgery Is Dependent on the Underlying Pathological Condition. Am J Sports Med. 2019 Jan;47(1):241-7. DOI: 10.1177/0363546517749213.
70. Koorevaar RC, van 't Riet E, Gerritsen MJ, Madden K, Bulstra SK. The Influence of Preoperative and Postoperative Psychological Symptoms on Clinical Outcome after Shoulder Surgery: A Prospective Longitudinal Cohort Study. PloS one. 2016;11(11):e0166555. DOI: 10.1371/journal.pone.0166555.
71. Hartel BP, Alta TD, Sewnath ME, Willems WJ. Difference in clinical outcome between total shoulder arthroplasty and reverse shoulder arthroplasty used in hemiarthroplasty revision surgery. Int J Shoulder Surg. 2015 Jul-Sep;9(3):69-73. DOI: 10.4103/0973-6042.161426.
72. Sallay PI, Reed L. The measurement of normative American Shoulder and Elbow Surgeons scores. J Shoulder Elbow Surg. 2003 Nov-Dec;12(6):622-7. DOI: 10.1016/S105827460300209X.
73. Hobby JL. Commentary on Asshein and Finsen. The DASH and the QuickDASH instruments. Normative values in the general population in Norway: and Iwatsuki et al. Comparative responsiveness of the Hand 20 and the DASH-JSSH to clinical changes after carpal tunnel release. J Hand Surg Eur Vol. 2014 Feb;39(2):152-4. DOI: 10.1177/1753193413493733.

74. Aasheim T, Finsen V. The DASH and the QuickDASH instruments. Normative values in the general population in Norway. J Hand Surg Eur Vol. 2014 Feb;39(2):140-4. DOI: 10.1177/1753193413481302.
75. Klum M, Wolf MB, Hahn P, Leclere FM, Bruckner T, Unglaub F. Normative data on wrist function. J Hand Surg Am. 2012 Oct;37(10):2050-60. DOI: 10.1016/j.jhsa.2012.06.031.
76. Jester A, Harth A, Wind G, Germann G, Sauerbier M. Disabilities of the arm, shoulder and hand (DASH) questionnaire: Determining functional activity profiles in patients with upper extremity disorders. J Hand Surg Br. 2005 Feb;30(1):23-8. DOI: 10.1016/j.jhsb.2004.08.008.
77. Hunsaker FG, Cioffi DA, Amadio PC, Wright JG, Caughlin B. The American academy of orthopaedic surgeons outcomes instruments: normative values from the general population. J Bone Joint Surg Am. 2002 Feb;84-A(2):208-15.
78. 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 Mar;41(3):582-9. DOI: 10.1177/0363546512472330.