6-14-2019 9:30 AM

Measuring Quality of Life in patients with Rotator Cuff Disorders

Rochelle Furtado, *The University of Western Ontario*

Supervisor: MacDermid, Joy C., *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Master of Science degree in Health and Rehabilitation Sciences

Abstract

Rotator cuff disorders (RCDs) are a common musculoskeletal problem with a higher prevalence in women and after the age of 60. Due to the increasing need for patient-centered care, researchers have been directed towards the use of patient reported outcomes (PROs) to evaluate the progress of patient recovery, as the experience varies across individuals.

Recently, the Western Ontario Rotator Cuff Index was reduced in items to create the shortened version of the Western Ontario Rotator Cuff Index (Short-WORC). However, the Short-WORC's psychometrics have yet to be evaluated. Therefore, this mixed-methods thesis aimed to evaluate the psychometric properties of the Short-WORC in a patient population of RCDs. First, we created the framework of evaluation by synthesizing the literature on the cultural adaptions of the WORC. Then, we quantitatively assessed the reliability and validity, and qualitatively assessed the content validity of the Short-WORC in a prospective rotator cuff population.

Overall, the findings suggest that the Short-WORC demonstrates evidence of validity, reliability and sensitivity to change when measuring quality of life in rotator cuff pathology. Findings suggest the appropriate recommendations and advancement of clinical research in rotator cuff pathology.

KEYWORDS: rotator cuff disorders, quality of life, musculoskeletal disorders, patient reported outcomes, Short-WORC, WORC, psychometric properties

Co-Authorship Statement

Research question, specific objectives and individual study design were developed by Rochelle Furtado and Joy C. MacDermid with inputs from Dianne Bryant, Kenneth J. Faber and George S. Athwal. Co- investigators were recruited when additional raters with specific expertise were required. Thesis advisory committee members were included as co-authors for specific chapters based on their input to individual chapters in this thesis work. The authors and specific roles for each chapter of the thesis are listed below:

CHAPTER 1: Introduction

Rochelle Furtado – responsible for study design, literature review, quality appraisal, data extraction, narrative synthesis and manuscript writing
Joy C. MacDermid – study design, data analysis and reviewed manuscript
Dianne Bryant - Study design and reviewed manuscript
Kenneth J. Faber – Study design, provided subjects and reviewed manuscript
George S. Athwal – Study design, provided subjects and reviewed manuscript
Goris Nazari – Quality appraisal, data extraction and reviewed manuscript

CHAPTER 2: Evaluating the reproducibility of the short version of the Western Ontario Rotator Cuff Index (Short-WORC) prospectively

Rochelle Furtado – Primary author, study design, data analysis and wrote manuscript
Joy C. MacDermid – Study design, data analysis and reviewed manuscript
Dianne Bryant - Study design and reviewed manuscript
Kenneth J. Faber – Study design, provided subjects and reviewed manuscript
George S. Athwal – Study design, provided subjects and reviewed manuscript

CHAPTER 3: Evaluating the validity of the Short version of the Western Ontario Rotator Cuff Index (Short-WORC) prospectively

Rochelle Furtado – Primary author, study design, data analysis and wrote manuscript
Joy C. MacDermid – Study design, data analysis and reviewed manuscript
Dianne Bryant - Study design and reviewed manuscript

Kenneth J. Faber – Study design, provided subjects and reviewed manuscript
George S. Athwal – Study design, provided subjects and reviewed manuscript

CHAPTER 4: Interpretation and content validity of the items of the numeric rating version Short-WORC to evaluate outcomes in management of rotator cuff pathology: A cognitive interview approach

Rochelle Furtado – Primary author, study design, data analysis and wrote manuscript
Joy C. MacDermid – Study design, data analysis and reviewed manuscript
Dianne Bryant - Study design and reviewed manuscript
Kenneth J. Faber – Study design, provided subjects and reviewed manuscript
George S. Athwal – Study design, provided subjects and reviewed manuscript

CHAPTER 5: General Discussion

Rochelle Furtado – sole author

Acknowledgments

I would like to thank my supervisor Dr. Joy C. MacDermid for her continual support, guidance and patience through these crazy two years. Thank you for being a role model and mentor through both my academic and personal life. I admire your passion, knowledge and dedication to both the field of MSK and all your students. Here's hoping that the next 5 years sparks more exciting and innovating research together, because clearly, you're not getting rid of me that easily.

I would like to thank my advisory committee members: Dr. Kenneth Faber, Dr. Dianne Bryant and Dr. George Athwal. Your advice and guidance was more than appreciated. Despite your busy schedules, you always made time to give my work endless feedback. Thank you for pushing my research career. I am l forward to continuously collaborating with you all in the future!

To my lab mates at HULC, thank you for your support and collaborations through the last two years! It has been a pleasure working with you all, and your support will always be appreciated! Special thanks to Christina Ziebart, Pavlos Babos and Goris Nazari. You guys are some of the most dedicated and hardworking researchers I have ever seen, and I hope one day I can be at your crazy level.

Special thanks to Steve Lu and Katrina Munro for all their help through recruitment and teaching me the ins and outs of HULC. Kat, thanks for all your support in the last two years, lucky for you, I still have 5 more years to continue to bug you. I would also like to thank all the surgeons, fellows, residents and nurses at the Roth | McFarlane Hand and Upper Limb Clinic for their help in screening and recruiting patients for the studies.

Special thanks to all the wonderful HRS friends I have made here at Western. Grad school is never easy, but I am glad that I got to share some of the best memories, with some of the best people I could have ever met.

Lastly, thank you to my parents and brother who I dedicate this thesis to. You guys drove me crazier than writing this thesis did, but I could not have done it without you. Thanks for making me the person I am today, and I hope to always make you proud.

Table of Contents

List of Tables

List of Figures

CHAPTER 1: Introduction

1.1 Rotator cuff disorders and treatments

The rotator cuff (RC) is comprised of four muscles that help stabilize the glenohumeral joint. The supraspinatus, infraspinatus, teres minor and subscapularis muscles provide dynamic stabilization of the glenohumeral joint. The RC provides the power for shoulder abduction and at least 90% of the power in external rotation. While the RC is an essential component for normal shoulder movement, it is also the leading site of many musculoskeletal injuries. [1-3]

Furthermore, the range of rotator cuff injuries vary amongst age groups and occupation. The prevalence of RCD, symptomatic or asymptomatic, increase with age from 9.7 percent under the age 20, to 62% over the age of 80.[23] RCDs lead to a large source of morbidity amongst manual laborers and those partaking in repetitive movements. [23] In young adults, rotator cuff injuries arise from shoulder instability, but older adults commonly face rotator cuff tears. These tears result in either partial (incomplete tear that doesn't extend through the entire thickness of the tendon) or complete tears (tear that extends through the complete thickness of the tendon).[1,2] Acute tears usually lead to shoulder dislocations or avulsion injuries, commonly in younger adults. The tears are heavily age dependent and considered a degenerative process. On the other hand, impingement, is another common rotator cuff disorder, and results from decreased space beneath the anterior acromion and the acromioclavicular joint. This can arise from problems with the balance of the dynamic and static stabilizers of the shoulders, as commonly seen in middle-age adults.[3]

Additionally, we see RCDs having a higher incidence rate in females, than males (90 cases per 100 000 person-years in women and 83 per 100 000 person-years in men; $p<0.001$). When comparing incidence rates between males versus females, there is no reported peak incidence between the age group of 55-59 years. [24] However, it is reported that females from 40-54 years old have a higher incidence rate than males.[24] Furthermore, other studies have shown rotator cuff disorders to be more prevalent in the working female population (9%) than male working population (6.8%). [25,26]

As the spectrum of rotator cuff disorders (RCD) vary, so do the approaches of treatments and medical management.[3] One of the most common forms of non- operative rotator cuff repair

(RCR) is through rehabilitation. Rehabilitation of the RC requires a basic understanding of the surgical procedures and expectation of outcome by the clinician. Rehabilitation programs are based on a patient's age, health status, compliance and injury. Rehabilitation focuses on restoring range of motion and restoring patient quality of life through exercises.[1-3] Exercises administered by physical therapists aim to allow patients to return to similar levels of previous function, through stressing stability and avoiding excessive mobility. The goals of rehabilitation include initial pain management and secondary muscle strengthening that will reposition the humeral head within the glenoid. As shown in previous literature, asymptotic tears should be managed non-operatively and the newly diagnosed tears should focus on core strengthening before surgical repair. Kuhn et al., have furthered shown that after a two year RCT, approximately 75% of patients who had only physical therapy to treat their full-thickness tear, have had a more effective recovery.[27] Additionally, another form of non-operative management includes corticosteroid injections.[4] The subacromial injections of local anesthetic and cortisone help to combat pain control, but does not necessarily improve healing. Clinicians may administer these on average about 3 times a year, but it is often seen in combination with rehabilitation exercises. [3] The other form of RCR, involves surgical treatment which aims to restore the continuity of the tendon and relieve pain. However, the outcomes of surgery can be difficult to predict, as the surgery depends on the quality of the tendon substance, the length of the time since the injury has occurred, and the underlying pathology that brought upon the injury. [1]

1.2 Quality of Life (QoL)

One of the main goals of RCR is to restore quality of life (QoL) for a patient. According to the World Health Organization Quality of Life Group (WHO-QOL), quality of life is defined as the individual's perceptions of their position in life in the context of the culture and value systems in which they live, in relation to their goals, expectations, standards and concerns.[5] QoL further includes the perceptions about life satisfaction/enjoyment, and is overall subjective to the individual.[5] RCDs are shown to hinder the QoL of an individual, therefore, the aim is to restore QoL with rehabilitation. Since QoL is an outcome of RCR, researchers aim to measure this outcome when determining a successful RCR. One of the most common ways to measure QoL is through patient reported outcome measures (PRO).[6,7]

1.3 Shoulder Patient Reported Outcomes (PROs)

Patient reported outcomes (PROs) provide subjective information to help the clinician assess physical function, psychosocial issues or general health related quality of life.[7] Usually administered as a questionnaire, a PROs aims to capture the possible effect of the condition, disease or intervention by incorporating the experience and perception of the patient.[7] Currently, there are many shoulder specific PROs for clinicians to use, making it difficult to choose the most appropriate one. To aid with communication and decision-making, PROs are classified as either disease-specific (rotator cuff specific) or population-specific.

1.4 Western Ontario Rotator Cuff Index (WORC)

As researchers shifted their focus towards measuring both the disability and QoL that arises from RCDs, the development of PROs increased such as the Western Ontario Rotator Cuff Index (WORC). The WORC by Kirkley, Alvarez and Griffin[8] is a disease-specific PRO that measures QoL and is categorized into the 5 domains of Physical symptoms, Sports/Recreation, Work, Lifestyle and Emotion in 21 items. Each item has a possible score from 0–100 (100 mm Visual Analogue Scale). Scores can be computed for individual subscales and summated for a total score, which can range from 0–2100, with a higher score representing lower QoL. Each subscale is scored and summed to a percentage reported by subtracting the total from 2100, dividing by 2100, and multiplying by 100 (physical symptoms/600, sports and recreation/400, work/400, lifestyle/400, and emotions/400).[8]

The goal of the WORC was to overcome the challenges that current shoulder PROs encountered. Deficiencies such as using too many items, having only parts of the tool that are used to generate a score, leaving patients out of the process of item selection and using double-barreled questions.[6,7,9] When evaluating the WORC in a RCD population, the WORC demonstrated strong test-retest reliability across several studies (ICCs 0.84 to 0.96). The construct validity of the WORC was determined by comparing other disability instruments such as the American Shoulder and Elbow Surgeons score (ASES) ($r = 0.68$) and the Disabilities of the Arm, Shoulder and Hand (DASH) ($r = 0.63$), demonstrating evidence that the scores generated by the WORC change similarly to these other validated measures of the same construct.[8] While the WORC demonstrates

evidence of being a reliable and valid tool for measuring QoL in a RCD population, there were some limitations. For example, some have reported challenges when administering a visual analogue scale, the lack of validation of the individual subscales and overall response burden for patients when completing the WORC. [8,10]

1.5 Shortened Western Ontario Rotator Cuff index (Short-WORC)

Recently, a shorter version of the Western Ontario Rotator Cuff Index (Short-WORC) was created with an abbreviated scale and a single score summary.[10] A confirmatory analysis was employed to reduce the original WORC to the 7-item Short-WORC. The Short-WORC contains items from the domains of work and lifestyle only. Therefore, the Short-WORC mainly focuses on the activity limitations that arise from rotator cuff disorders, rather than QoL.[9,10] Compared with the WORC, the Short-WORC eliminates administrative burden and has psychometric equivalence when used to evaluate rotator cuff pathology. However, current studies assessing the Short-WORC were conducted by extracting Short-WORC data from existing datasets of the original WORC. [9,10] Future studies need to administer the Short-WORC on a prospective patient population, in order to evaluate the validity, reliability and responsiveness in samples that represent the spectrum of rotator cuff disorders.

1.6 Criteria for evaluation

According to the International Society for Quality of Life Research (ISOQOL),[11] the minimum standard following the development of a new PRO is to provide strong evidence of reliability, validity (content validity, construct validity, responsiveness); interpretability of scores; quality translation, and acceptable patient and investigator burden. A PRO that possesses these traits is more useful for both clinicians and researchers.[12]

Reliability is the degree to which a PRO is free of measurement error and can be further evaluated as 1) test-retest reliability (stability) and 2) internal consistency (homogeneity). Test – retest evaluates the ability of a PRO to provide consistent scores over time in a stable population.[13,14] It is usually measured by calculating the intra class coefficient (ICC). In contrast, the internal consistency measures the relatedness of the items in a PRO, and is calculated by Cronbach's alpha. A strong PRO will usually score between 0.7-0.9.[14]

Validity can be defined by the extent to which the PRO measures the construct it is intended to measure.[15] Validity can be further broken down into the components of content validity, construct validity and criterion validity. Content validity is the extent to which the PRO represents the most relevant and important aspects of a concept in a given context. Content validity can be evaluated by both qualitative or quantitative methodologies.[16] Qualitative methodologies such as qualitative description or interpretative description can be used in the development of PRO to support content validity or in the evaluation of the content validity of existing measures. Cognitive interviews aim to evaluate the clarity of the instructions, the perceived content of each item on the PRO and if the intended meaning of each item is easily and correctly interpreted by participants.[17] This process includes evaluation of instructions, recall periods and response scales. This method uses a talk aloud technique of probes that allow participants to verbally express their thoughts and responses to each item on the PRO. [16,17]

Another type of validity, construct validity; is defined by how the scores on the PROs relate to other PROs evaluating similar constructs.[18] This can be further confirmed by measuring if the PRO is consistent with theoretically derived a priori hypotheses concerning the concepts being evaluated. [15] Usually, construct validity will have predefined hypotheses on the expected associations among measures that are similar to the measured PRO. This can be calculated through Pearson's correlation coefficient (r), which defines that a moderate to large correlation (> 0.40) between constructs are expected to support (concurrent convergent) validity of the measure.[11] Another type of validity, criterion validity; is assessed by testing a new PRO against an independent criterion or standard (concurrent validity) or against a future standard (predictive validity).[19] Criterion validity is an estimate of the extent to which a measure agrees with a current gold standard.[19]

The measurement property of responsiveness, is defined as the PRO's ability to capture changes over time in the construct being evaluated.[13] A measure that is not responsive cannot capture the change in a patient's health status resulting from rehabilitation. [20,21] Therefore, to evaluate a clinical change we measure responsiveness by either the anchor-based or distribution-based method. In the anchor-based method, another outcome measure is used as a criterion for change, such as the Global Rating of Change scale (GRC). Then, a minimal clinically important difference (MCID) is calculated, to indicate the change in health conditions as a response to the

treatment administered. While the method is widely used, criticisms of the GRC can potentially compromise this method. In contrast, the distribution based method calculates the effect sizes (ES) and standardized response means (SRMs) based on change scores and variability.[20] This method is based on the obtained sample characteristics (SD) and not as clearly linked to an important reference. Different methods of calculations may result in different MCID values and vary based on sample size. Sensitivity to change (longitudinal validity) is the "ability of an instrument to measure change in the state regardless of whether it is relevant or meaningful to the decision maker". As shown by Liang et al., ES and SRM can be used as coefficients to evaluate whether a measure is able to detect change.[22]

PROs should always be evaluated for their psychometric properties before clinical use.[7] Regardless if a PRO is validated in its original context, once adapted for another culture, we must re-evaluate its measurement properties. Therefore, researchers can use specific guidelines in order to evaluate culturally adapted PROs, to verify that they have retained the intended construct after adaption.[10] We aimed to evaluate the parent WORC, as it has been previously adapted for 10 different languages. Our goal was to synthesize the literature and analyze how different cultures evaluated different psychometric properties for the WORC. This would be the groundwork for understanding how we should evaluate the psychometric properties of the Short-WORC in a prospective cohort.

Cross-cultural adaptions and measurement properties of the WORC (Western Ontario Rotator Cuff Index): a systematic review

Abstract

Background: To evaluate the translations, cross-cultural adaptation procedures and measurement properties of the Western Ontario Rotator Cuff Index (WORC), when it is adapted for different cultures.

Methods: A systematic review was performed, considering different cultural adaptions of the WORC accessible through MEDLINE, CINHAL, EMBASE and/or Google Scholar. Included were prospective cohort studies that used an adapted version of the WORC to measure QoL in patients with rotator cuff disorders. All studies were evaluated according to the current guidelines for cross-cultural adaptations and measurement properties. Data extraction and ratings of adherence to the guidelines were conducted by two independent reviewers.

Results: The search retrieved 13 studies that met the inclusion criteria. According to the recommended guidelines for cross-cultural adaptations, 8 studies performed 100% of the steps, 2 studies performed 80% of the steps and 3 studies used previously translated measures. When evaluating the studies' psychometric properties based on the quality criteria, none of the studies reported all recommended measurement properties. 100% of studies reported the measurement property of reproducibility (reliability), but none of the studies reported reproducibility (agreement). Internal consistency was fully reported according by 15% of studies. 30% of studies reported construct validity. Overall, the study by St-Pierre et al. was the most successful in fully reporting 100% of the cross-cultural adaption guidelines and 83% of the quality criteria.

Conclusions: Although the majority of studies demonstrated proper adaptation procedures, testing of the measurement properties were inadequate. It is recommended that the current adapted versions of the WORC undergo further testing before use in clinical practise, and researchers continue to adapt the WORC for different cultures as it proves to be an appropriate instrument for assessing rotator cuff pathology.

Key words: rotator cuff disorders; translation; psychometric properties; WORC; quality of life; patient reported outcomes; shoulder; rotator cuff tear

1.1

A version of this work has been accepted for journal publication: Furtado R, MacDermid JC Nazari, G., Bryant DM, Faber K, and Athwal G Cross-cultural adaptions and measurement properties of the WORC (Western Ontario Rotator Cuff Index): a systematic review. Health and Quality of Life Outcomes. (2019)

Introduction

Shoulder pain is one of the most commonly reported musculoskeletal problems that result in the restriction of work and/or social activities.[1-3] Rotator cuff disorders (RCDs) are the most common causes of shoulder pain, as chronic tendon degeneration of the cuff results in a loss of tendon integrity that ranges from partial to massive tears.[3] RCDs are highly prevalent in males, and more frequent in working individuals over the age of 60.[2,3] Overall, untreated RCDs eventually lead to the loss of quality of life

(QoL).[1-3]

Measuring QoL can help to determine prognosis and evaluate treatment outcomes in patients with RCDs.[2-4] In order to estimate QoL, self-reporting through patient reported outcomes (PROs)[1-4] is required. The Western Ontario Rotator Cuff Index (WORC), developed by Kirkley et al is one of the most validated disease-specific questionnaires to measure QoL in patients with RCD.[5] The WORC focuses on 5 domains; 1) pain and physical symptoms, 2) sports and recreation, 3) work, 4) lifestyle, and 5) emotions. The WORC has 21 items that respondents answer on a visual analogue scale, with anchors of "no pain/difficulty and extreme pain/difficulty". Items chosen for the WORC were derived from a variety of published health status scales, discussions with healthcare professionals, and interviews with a variety of patients with rotator cuff pathology.[4-7]

While there are a variety of PROs for evaluating and detecting changes in a patient's clinical condition over time, most were developed in English.[6-8] Due to the increasing globalization and importance of using these tools across cultures, researchers have been directed towards the translation of these outcome measures.[6,7] The availability of PROs for different cultures is not only economical but can facilitate future comparisons among different populations; as long as the translated equivalent is successful.[8] Therefore, PROs need to be accurately translated, cross–culturally adapted and assessed for their psychometric measurement properties.[7,8]

For an adapted measure to be applied to the intended population, careful attention to word change and question structure is required.[6-8] The cross-cultural adaption process, verifies the equivalence with the original version and resolves any cultural or health differences amongst countries.[6,9] Additionally, it is also important to evaluate the psychometric properties of the adapted

measure.[9,10] Evaluation after translation can verify if the adapted measure retains the psychometric properties of the original, as discrepancies between cultures can influence the results.[6,8-10] Therefore, guidelines have been developed to help researchers critically analyze these studies.[6,10-12]

Although the WORC has strong psychometric properties [1,2,13] in an English context, there is a concern regarding the cross-cultural adaptation procedures and measurement properties when translated. Therefore, this systematic review aims to evaluate the translations, cross-cultural adaptation procedures and measurement properties of the WORC, when adapted for different cultures.

Methods

Study Selection

We conducted a systematic review of studies that addressed the translation process and psychometric testing of the WORC in different cultures. The systematic searches were performed in the following key electronic databases: MEDLINE (Ovid), EMBASE, EBSCO- Host (CINAHL), and Google Scholar. Search terms and Boolean operators (AND or OR) used were: Western Ontario Rotator Cuff Index AND validation OR translation OR cross-cultural adaption AND different languages (e.g., German). The searches were not limited by language or publication date. The final search was May 13, 2018.

Inclusion Criteria

Studies were considered eligible for inclusion if they assessed a cross-cultural adaption of the WORC and its measurement properties in a specific language. Studies must be published as a full manuscript in a peer – reviewed journal. Thesis/dissertations, books and abstracts from conferences were excluded. There were no language restrictions.

Data Extraction and Analysis

Demographics of each study were extracted to include information on patient age, sex, and pathology. Data regarding the translation and cross-cultural adaptation were extracted to assess each design. The translation methods for each study were classified according to the Guidelines for the

process of Cross-Cultural Adaption of Self-report Measures[11]. These cross-cultural adaption guidelines state an accurate translation must include an initial translation, synthesis of translations, back-translations, reviews by the expert committee and the pre-test version of the instrument. We also extracted data relating to the measurement properties of each study. These measurement properties were evaluated according to the Quality Criteria for Measurement Properties of Health Status Questionnaires[10]. This quality criteria evaluates: construct validity, internal consistency, reproducibility (agreement and reliability), agreement, responsiveness and ceiling and floor effects. Other measurement properties such as content validity and interpretability are only relevant to the development of original questionnaires, and therefore, not relevant to the scope of this review. Additionally, item criterion validity is measured when there is a gold standard of criteria available for comparison. [6] Shoulder assessments do not have a gold standard criteria for item selection, therefore, this property was excluded from the review. Tables were used to describe both the quality of testing and clinimetric results. This approach has been frequently used in a variety of systematic reviews for health–related questionnaires.[6-8]

Data extraction and ratings were performed by the first author (R.F.) and then reviewed by an independent reviewer (G.N.). Any disagreements between the rater and independent reviewer were discussed to reach a consensus.

Results

From the search strategies, 114 studies were retrieved but only 13 met eligibility criteria. The 13 versions represent 10 different languages/cultures; Chinese[14], Dutch[15,16,17], French-Canadian[13], Japanese[18], Norwegian[19], Persian[20], Portuguese-Brazilian [21,22], Spanish[23], Swedish[24] and Turkish[25]. There was more than one study reporting clinimetric testing of the Dutch[15-17] and Portuguese-Brazilian[21,22]. All Dutch versions were conducted independently; Wiertsema et al reported on the reproducibility and translations of the WORC[15], Wessel et al reported on the reliability, reproducibility and cognitive interviewing of creating a conceptually equivalent version[16] and de Witte et al reported on the reliability and responsiveness of the WORC[17]. The Portuguese-Brazilian versions were conducted by the same group of researchers, however, one study focused on only the cross-cultural adaption process[22] and the other study focused on the evaluation of the psychometric properties[21].

Table 1 shows the demographic characteristics of the respective populations tested in the 13 studies. All studies included both male and female participants. While the literature recommends a minimum sample size of 100 patients, there are some exceptions. [35] For example, when evaluating content validity with qualitative methods, a sample size under 100 is justified.[35] In this review, all studies except the Portuguese –Brazilian[21] study (n=30) had more than 50 patients. Patients were treated for a partial or a full rotator cuff tear, tendinopathy, impingement syndrome or calcific tendonitis.

Table 2 describes the ratings of the cross-cultural adaptions according to the Guidelines for the Process of Cross-Cultural Adaptions of Self-Report measures[11]. From the 13 eligible studies, 10 studies performed 100% of all the recommended cross-cultural adaption guidelines when performing the initial step of translation.[13-16,18-20,22,23,25] These 10 studies also performed 100% of all recommended cross-cultural adaptation guidelines for the step of synthesis.[13-16,18-20,22,23,25] 9 studies performed 100% of the back-translation step according to the cross-cultural adaption guidelines.[13-16,18-20,23,25] The Portuguese-Brazilian[22] study performed 50% of the back-translation step according to cross-cultural adaptation guidelines, as they did not have two translators in the process. 10 studies performed 100% of the cross-cultural adaption guidelines for the step of expert committee review. [13-16,18-20,22,23,25] 9 studies performed 100% of the cross-cultural adaption guidelines for the step of pre-testing.[13-16,19,20,22,23,25] The Japanese[18] study performed 50% of the cross-cultural adaption guidelines for the step of pre-testing, as they did not provide the sample size used for pilot testing their questionnaire. 3 studies[17,21,24] used pre-translated versions of their questionnaires and therefore, did not report the translation process. 12 out of 13 studies[13-25] used translation guidelines proposed by Guillemin, Bombardier and Beaton[11]. While, the Turkish[25] study referred to the guidelines by Acquadro C, Jambon B, Ellis D, and Marquis P. [27]

Table 3 presents the ratings of the evaluated measurement properties according to the Quality Criteria for Measurement Properties of Health Status Questionnaire[10] for each study. Overall, 12 studies evaluated the measurement property of reliability.[13-20,22-25] These 12 studies followed 100% of the quality criteria for measuring reliability; using test re-test and Cronbach's alpha respectively. The measurement property of agreement was not adequately evaluated in any of the studies. 62% of studies[13-16,18,19,24,25] were rated as having a doubtful design, as only 50% of the quality criteria was followed. This meant that these studies had designs where the minimal

important change (MIC) was not defined and there were no convincing arguments that stated agreement to be acceptable. These studies reported agreement through standard error of the mean (SEM) or minimal detectable change (MDC) values, instead of MIC values. 30% of studies[17,20,21,23] did not provide any information or evaluate the measurement property of agreement in their study. Only the French-Canadian and Swedish studies[13,23] followed 100% of the quality criteria when measuring the property of internal consistency. 10 studies[14-17,18-21,24,25] performed 50% of the steps according to the quality criteria, as they did not include a factor analysis. Only the French-Canadian study[13] was able to follow 100% of the quality criteria when evaluating the measurement property of responsiveness. 4 studies[14,18,19,24] followed only 50% of the recommended quality criteria to evaluate the property of responsiveness. These studies had designs in which the smallest detectable change group was bigger than the MIC OR the MIC and/or limits of agreement (LOA) were less than 1.96. Furthermore, 5 studies did not report the measurement property of responsiveness. 5 studies followed 100% of the quality criteria steps when evaluating construct validity.[13,14,18,19,23] 7 studies did not evaluate or report the measurement property of construct validity.[15,16,17,20,21,24,25] The Chinese[14], Norwegian[19] , Swedish[24], Dutch[16] and French-Canadian[13] studies followed 100% of the quality criteria for assessing the measurement property of ceiling or floor effects. The Persian study[20] followed 50% of the quality criteria when measuring ceiling and floor effects, as more than 15% of the respondents achieved the highest or lowest possible scores, despite having an adequate design and method. 54% of studies did not report any floor or ceiling effects. [15,18,21-23,25]

Discussion

This systematic review evaluated the cross-cultural adaption procedures and measurement properties reported in 13 adapted versions of the WORC[13-25]. Overall, the key findings of this review demonstrate that regardless of adaption methods, there is a lack of clinimetric testing in the majority of translated versions. Therefore, further validation of these adapted measures is needed to ensure they are able to measure the intended construct.

The primary outcome of the WORC is to evaluate disability related to RCDs and its effects on health-related quality of life.[5] Therefore, the intended patient population includes acute rotator cuff tendinitis, rotator cuff tendinosis with no tear, partial and full thickness tears and rotator cuff tear arthropathy.[5] While the majority of studies in this review recruited from this spectrum, some

studies included calcific tendonitis.[15-17] It is important to highlight that calcific tendonitis does not fall under the scope of rotator cuff pathology, as it occurs from cell-mediated calcification inside the tendon. This can lead to patients experiencing extreme symptoms of pain and impingement, therefore, being confused with rotator cuff tear or impingement syndrome. [30] While the co-existence of calcific tendonitis with rotator cuff tear is not uncommon, calcific tendonitis is a non-degenerative condition that does not result in the tendon becoming torn or pathologic.[30,31] Since the WORC is specific to rotator cuff pathology, inclusion of these patients hinder the homogeneity of the sample. Therefore, researchers should always recruit study populations that preserve the intended meaning of the outcome measure.[32]

One issue that made the ratings less certain, was the lack of detail provided for the cross-cultural adaption processes used in the individual studies. Five studies[16,18,20,22,24] in this review provided a brief explanation of the translation processes. The Dutch[16] and Portuguese-Brazilian studies[22] assessed content validity by using cognitive interviewing. The results from the interviews demonstrated that the adapted WORC was only a reliable measure for patients, once cultural modifications had been applied to the individual items. Therefore, it is highly recommended to provide all relevant details of the translation process and discuss all issues that may have occurred, so that future researchers can anticipate when translating. In order to ensure items fit the context of the culture, many researchers will change individual words or sentence structure. For example, the Chinese study[14] noted issues with translations of item 17. As most families in China are traditional, the term "rough-housing or horsing around" is inapplicable and had to be modified to the Chinese culture. Therefore, while researchers modify items that do not fit the context or culture of the target population, it must be done carefully to ensure that content validity is retained.

The back-translation step is often overlooked, but is critical according to the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) 's guidelines.[33] Currently there is little agreement on how the back translation should be performed, but one of the translators should be of the origin language. This is to limit the amount of words or phrases that may not respect the speech patterns or colloquialisms of the target culture. For example, since there are a variety of dialects in Portuguese, the Portuguese-Brazilian version would have to be translated again to be used in Europe. ISPOR guidelines recommend that health-related PROs use conceptual translations,

as they deal with subjective terms.[33] Therefore, researchers should adapt accordingly to maintain the intended meaning of the construct.[15,16,26]

Reliability was evaluated in all studies and performed correctly according to the quality criteria. All studies in this review reported an Intraclass Correlation Coefficient (ICC) value of over 0.70, which the quality criteria rates as excellent.[11] However, only the French–Canadian[13], Japanese[18] and Dutch[15] studies provided the type of ICC model and/or gave a description of the confidence interval used. Reporting the type of ICCs used is important to distinguish results that maybe under - or overestimated. According to the quality criteria, reliability established by McGraw and Wong is preferred as systematic differences are considered to be part of the measurement error.[11,34] The quality criteria also defines reliability by having an adequate measurement interval.[11] Therefore, a time period between the repeated administrations should be long enough to prevent recall, but short enough to ensure that clinical change has not occurred. Generally, 1 to 2 weeks is appropriate, but there could be reasons to choose otherwise.[11] Some studies [13,20,21] in this review had a time interval that was too long or not long enough. However, they were able to justify that due to participants starting rehabilitation immediately after their initial evaluation, researchers needed to either extend or shorten the time intervals to maintain consistency. Therefore, it is important for studies to describe and justify their time period to ensure that patients have not been changed on the construct that is being measured.[28]

Agreement is another important measurement property that further evaluates the degree of which repeated measures applied to patients provide similar answers. It is easier to clinically interpret than the property of reliability, and provides the absolute error of measurement.[11] In this review, no study was able to fully evaluate agreement according to the quality criteria. The quality criteria recommend that studies should determine the MIC value because distribution-based methods do not provide a good indication of the importance of the observed change; however, studies in this review only report MDC values.[6,11] Ideally, studies should test reproducibility by assessing both reliability (relative error of measure) and agreement (absolute error of measure). [6]

According to the quality criteria, responsiveness is a measure of longitudinal validity, and should be able to distinguish clinically important change from measurement error.[11] Responsiveness

was assessed by 6 studies[13,14,17-19,24] and only the French- Canadian[13] and Dutch [17] studies reported responsiveness at 100% according to the quality criteria. These studies were able to report MIC values that were greater than the SDC, which were consistent with Kirkley et al.[5] However, it is important to note that there is more than one way to evaluate responsiveness according to the quality criteria. The area under the receiver operating characteristics (ROC) curve (AUC), which measures the ability to distinguish patients who have and have not changed according to an external criterion, is also acceptable. An AUC value of at least 0.70 is considered to be adequate.[11] Therefore, researchers should always try to find a way to report the responsiveness in order to certify that the translated measures can detect patient improvement.

Ceiling and floor effects are another important measurement property according to the quality criteria.[11] Ceiling or floor effects are present if more than 15% of patients achieve the lowest or highest possible score, respectively. In this review, only 6 studies[13,14,16,19,20,24] reported testing for ceiling and floor effects. If ceiling or floor effects were present, content validity, reliability and responsiveness are all negatively impacted.[6-8] This indicates that the highest and lowest scores cannot be distinguished from each other, and changes cannot be measured in these patients. Therefore, reporting floor or ceiling effects verifies if the translated measures would fail to detect patient improvement or deterioration.[6]

Construct validity was performed according to quality criteria in only 5 studies.[13,14,18,19,23] These studies formulated hypotheses concerning the concepts measured. The most important feature of construct validity is to formulate hypotheses α priori, and to specify the direction of the expected correlation and its magnitude. Stating the hypothesis is crucial, otherwise the risk of bias is high, and it would be easier to develop an alternative explanation for the low correlations, than to admit that the construct validity has been compromised.[6,11]

This review demonstrates that there were many inconsistencies with some of the reported measurement properties in the various adaptions of the WORC. In the systematic review of the cross-cultural adaption and measurement properties of the McGill Pain Questionnaire[8], it was observed that many properties were either not evaluated or inappropriately measured. This was also similar to findings of a systematic review that looked at cross-cultural adaptions and measurement

properties of various shoulders outcomes in Portuguese.[6] The lack of appropriately testing these measures creates challenges for researchers and clinicians. The goal with adapting validated PROs is to achieve equivalence. Therefore, researchers must focus on maximizing both the linguistic, cultural and structural system of health-related measurements.[6] By developing culturally equivalent versions of these instruments, we can promote the exchange of information from studies across different cultures, without constantly having to create new PROs.[6-8] Therefore, following the proper guidelines for cross-cultural adaptations and for testing measurement properties is critical.

Based on the findings from this review, the French-Canadian study[13] had performed 5 out of 6 steps at 100% according the quality criteria, and 100% of all recommended steps of the cross-cultural adaption guidelines. However, just because a study received the highest number of positive ratings, does not necessarily mean it is the best outcome measure. Ratings depend on the availability of information and the quality of reporting on the assessment. For example, newer outcome measures may have many indeterminate ratings of measurement properties, as they are yet to be evaluated. Furthermore, it is important to note that there is no overall quality score with these guidelines[10,11], is often done in systematic reviews of randomized clinical trials. Having an overall quality score assumes that all measurement properties are equally important, which is not always true. A successful outcome measure requires a variety of different qualities with respect to reproducibility and responsiveness.[11] In particular, evaluative PROs such as the WORC, require a high level of agreement to be able to measure important changes, which was lacking in the present studies.[11]

Overall, limitations of this study lie within the inclusion criteria, as this review was limited to the use of peer-reviewed journal articles only. This excluded original versions of dissertations and theses with unpublished data regarding measurement properties. Additionally, only articles published in English were used for this review, and therefore the German[29] translated version of the WORC was excluded. This was to avoid discrepancies that could arise from inaccurate translations of German to English.

Conclusion

Researchers should follow recommended guidelines when trying to adapt questionnaires for different cultures. Further validation of the adapted versions of the WORC is required before use in clinical practice.

Table 1. Demographic and clinometric characteristics of the study populations from each study.

Study Country (Language)	Year	Sample size(n)	Mean (SD) age	%female	%male	Shoulder condition
China (Chinese)[14]	2017	152	47.3 (9.5)	44.4	55.6	RC disorders that needed arthroscopic surgery
Netherlands (Dutch) [15]	2013	52	54.2(9.7)	58	42	Partial or full thickness RC rupture, calcific tendonitis, or RC tendinopathy
Netherlands (Dutch) [16]	2013	57	53	47	53	RC tear, calcific tendonitis, impingement/tendinosis/ tendonitis
Netherlands (Dutch) [17]	2012	92	55(8.7)	53	47	RC tear, calcific tendonitis, impingement
Canada (French – Canadian) [13]	2015	87	49.7 (12.4)	34.5	65.6	Tendinopathy, full or partial thickness RC tear
Japan (Japanese) [18]	2013	75	63.4 (11.1)	43	57	Impingement syndrome, tendinopathy, partial or full thickness RC tear
Norway (Norwegian) [19]	2008	74	51(11)	64	36	Shoulder pain or full-thickness rotator cuff tear
Iran (Persian) [20]	2009	120	46.7 (15.4)	45.6	48.7	Rotator cuff tendonitis, rotator cuff tendinosis with no tear, partial tear or full-thickness tear

Poland (Polish)[29]	2018	69	55.5	29	71	Had to be operated for rotator cuff disorders
Brazil (Portuguese-Brazilian) [21]	2008	100	56.7 (10.8)	69	31	Tendinopathy, full or partial thickness RC tear
Brazil (Portuguese-Brazilian) [22]	2006	30	55.1 (10.8)	46.7	53.3	Tendinopathy, full or partial thickness RC tear
Spain (Spanish) [23]	2015	60	57(12.3)	44	56	Tendinopathy, full or partial thickness RC tear
Sweden (Swedish) [24]	2016	65	60	42	58	Surgery for subacromial pain condition or RC disorder
Turkey (Turkish) [25]	2006	72	54.9 (9.9)	75	25	Impingement syndrome, full or partial RC tears

Table 2. Cross-cultural adaptions of the WORC into different languages that used the translation-based approach related to the guidelines for the Process of Cross-Cultural Adaption of Self-Report Measures

Studies	Translation	Synthesis	Back translation	Expert committee review	Pretesting
China[14]	+	+	+	+	+
Dutch [15]	+	+	+	+	?
Dutch [16]	+	+	+	+	+
Dutch [17]	N/A	N/A	N/A	N/A	N/A
French – Canadian [13]	+	+	+	+	+
Japanese[18]	+	+	+	+	?
Norwegian [19]	+	+	+	+	+
Persian [20]	+	+	+	+	+
Polish[29]	N/A	N/A	N/A	N/A	N/A
Portuguese-Brazilian [21]	N/A	N/A	N/A	N/A	N/A
Portuguese-Brazilian[22]	+	+	?	+	+
Spanish [23]	+	+	+	+	+
Swedish [24]	N/A	N/A	N/A	N/A	N/A
Turkish [25]	+	+	+	+	+

N/A = not applicable – The cross-cultural adaptions was not performed, only the clinometric tests. Questionnaires used in these studies have been previously translated. += positive rating; - = negative rating; 0= no information available; ?=unclear

Table 3. Measurement properties of the WORC adapted into different languages related to the Quality Criteria for Measurement Properties of Health Status Questionnaires.

Studies	Reproducibility (Agreement)	Reproducibility (Reliability)	Internal Consistency	Responsiveness	Construct Validity	Ceiling and floor effects
China[14]	?	+	?	?	+	+
Dutch [15]	?	+	?	0	0	0
Dutch [16]	?	+	?	0	0	+
Dutch [17]	0	+	?	+	?	+
French – Canadian [13]	?	+	+	+	+	+
Japanese[18]	?	+	?	?	+	0
Norwegian [19]	?	+	?	?	+	+
Persian [20]	0	+	?	0	0	-
Polish[29]	0	+	?	+	+	+
Portuguese-Brazilian [21]	0	+	?	0	0	0
Portuguese-Brazilian[22]	0	0	0	0	0	0
Spanish [23]	0	+	+	0	+	0
Swedish [24]	?	+	?	-	0	+
Turkish [25]	?	+	?	0	0	0

N/A = not applicable – The cross-cultural adaptions was not performed, only the clinometric tests.

Questionnaires used in these studies have been previously translated. += positive rating; - = negative rating;

0= no information available; ?=unclear

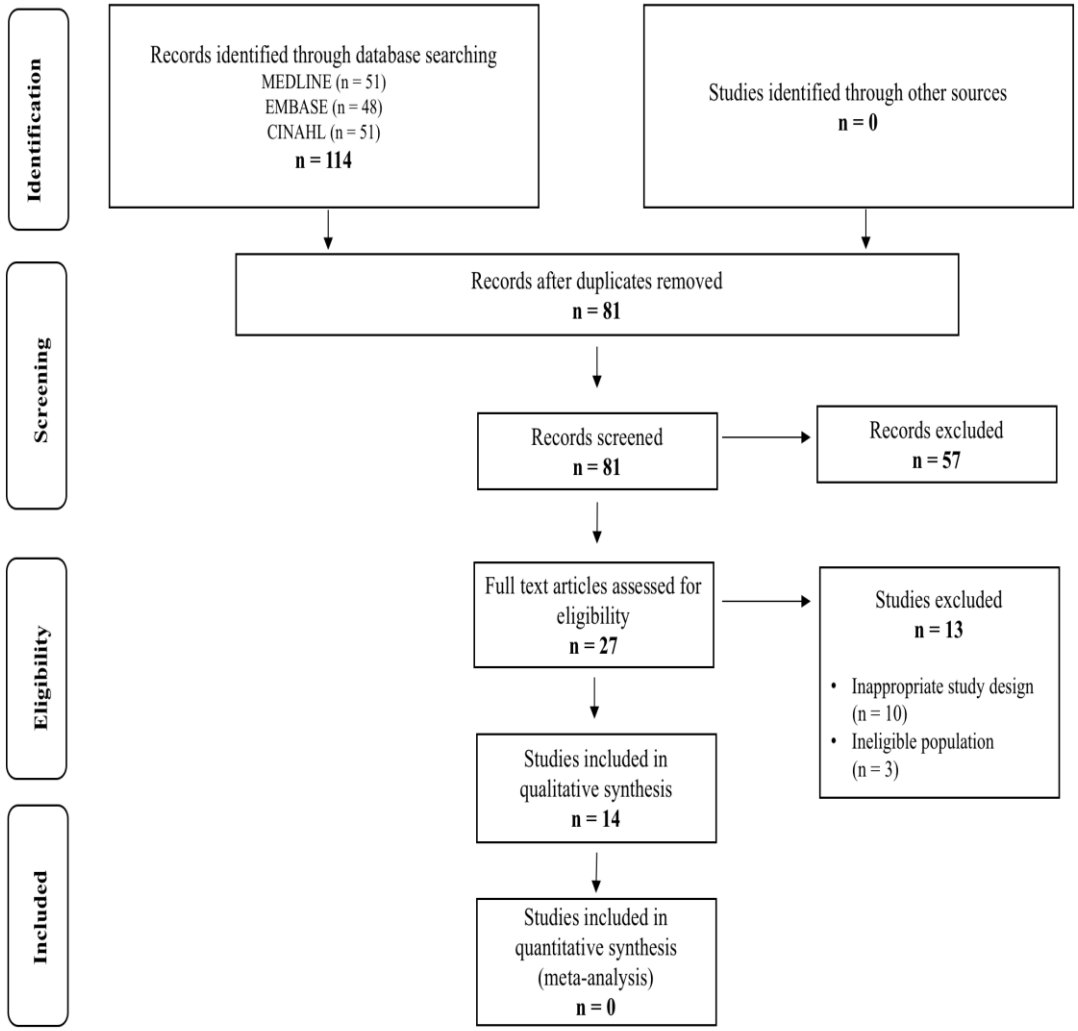

Figure 1. Flow diagram of literature search.

References

1. Dewan N, MacDermid JC, MacIntyre N. Validity and Responsiveness of the Short Version of the Western Ontario Rotator Cuff Index (Short-WORC) in Patients with rotator cuff repair. *J Orthop Sports Phys Ther*. 2018; 48:409-418. doi:10.2519/jospt.2018.7928

2. Dewan N, MacDermid JC, MacIntyre N, Grewal R. Reproducibility: reliability and agreement of short version of Western Ontario Rotator Cuff Index (Short-WORC) in patients with rotator cuff disorders. *J Hand Ther*. 2016; 29: 281– 291. doi.10.1016/j.jht.2015.11.007

3. Macdermid JC, Khadilkar L, Birmingham TB, Athwal GS. Validity of the QuickDASH in patients with shoulder-related disorders undergoing surgery. *J Orthop Sports Phys Ther*. 2015;45:25-36. doi: 10.2519/jospt.2015.5033.

4. Michener LA, Synder, AR. Evaluation of health-related quality of life in patients with shoulder pain: are we doing the best we can? *Clin Sports Med.* 2008; 27:491-505. doi:10.1016/j.csm.2008.03.001

5. Kirkley A, Griffin S. Development of disease-specific quality of life measurement tools. *Arthroscopy*. 2003; 19:1121-8. doi:10/1016/j.arthro.2003.10.028

6. Puga, Vanessa O. O., Lopes, Alexandre D., & Costa, Leonardo O. P. Assessment of cross-cultural adaptations and measurement properties of self-report outcome measures relevant to shoulder disability in Portuguese: a systematic review. Brazilian Journal of Physical Therapy. 2012;16:85-93. http://dx.doi.org/10.1590/S1413-35552012005000012

7. Gómez-Valero S, García-Pérez F, Flórez-García MT, Miangolarra-Page JC.Assessment of cross-cultural adaptations of patient-reported shoulder outcome measures in Spanish: a systematic review. *Shoulder Elbow*. 2017; 9:233-246. doi: 10.1177/1758573217694340

8. Costa LCM, Maher CG, McAuley JH, Costa LO. Systematic review of cross-cultural adaptations of McGill Pain Questionnaire reveals a paucity of clinimetric testing. *J Clin Epidemiol*. 2009;62:934-4. doi:10.1016/j.jclinepi.2009.03.019

9. Maher CG, Latimer J, Costa LOP. The relevance of cross-cultural adaptation and clinimetrics for physical therapy instruments. *Rev Bras Fisioter*. 2007;11:245-52. doi:10.1186/1471-2474-14-125

10. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007;60 :34-42. doi: 10.1016/j.jclinepi.2006.03.012

11. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine (Phila Pa 1976*). 2000;25:3186-91.

12. Maher CG, Latimer J, Costa LOP. The relevance of cross-cultural adaptation and clinimetrics for physical therapy instruments. *Rev Bras Fisioter*. 2007;11:245-52. doi: 10.1590/S1413-35552007000400002

13. St-Pierre C, Dionne CE, Desmeules F, Roy JS. Reliability, validity, and responsiveness of a Canadian French adaptation of the Western Ontario Rotator Cuff (WORC) index. *J Hand Ther*. 2015; 28:292-8 doi: 10.1016/j.jht.2015.02.001

14. Wang, W., Xie, Q., Jia, Z., Cui, L., Liu, D., Wang, C., & Zheng, W. Cross-cultural translation of the Western Ontario Cuff Index in Chinese and its validation in patients with rotator cuff disorders. *BMC Musculoskeletal Disorders*. 2017;18, 178. doi:/10.1186/s12891-017-1536-y

15. Wiertsema SH, Rietberg MB, Hekman KM, Schothorst M, Steultjens MP, Dekker J. Reproducibility of the Dutch version of the Western Ontario rotator cuff Index. *J Shoulder Elbow Surg.* 2013; 22:165-170. doi: 10.1016/j.jse.2012.03.014

16. Wessel, R. N., Wolterbeek, N., Fermont, A. J., van Mameren, H., Sonneveld, H., Griffin, S., & de Bie, R. A. The conceptually equivalent Dutch version of the Western Ontario Rotator Cuff Index (WORC)©. *BMC Musculoskeletal Disorders*. 2013; 14:362. doi:10.1186/1471-2474-14-362

17. de Witte, P. B., Henseler, J. F., Nagels, J., Vliet Vlieland, T. P., & Nelissen, R. G. The Western Ontario rotator cuff index in rotator cuff disease patients: a comprehensive reliability and responsiveness validation study. *The American Journal of Sports Medicine*. 2012;40:1611-1619. https://doi.org/10.1177/0363546512446591

18. Kawabata M, Miyata T, Nakai D, Sato M, Tatsuki H, Kashiwazaki Y, Saito H. Reproducibility and validity of the Japanese version of the Western Ontario Rotator Cuff Index. *J Orthop Sci*. 2013;18:705–11.doi.org/10.1007/s00776-013-0426-x

19. Ekeberg, O. M., Bautz-Holter, E., Tveitå, E. K., Keller, A., Juel, N. G., & Brox, J. I. Agreement, reliability and validity in 3 shoulder questionnaires in patients with rotator cuff disease. *BMC Musculoskeletal Disorders*. 2008; 9: 68. doi: 10.1186/1471-2474-9-68.

20. Mousavi SJ, Hadian MR, Abedi M, Montazeri A. Translation and validation study of the Persian version of the Western Ontario Rotator Cuff Index. *Clin Rheumatol*. 2009;28:293-299. doi: 10.1007/s10067-008-1042-6

21.Lopes AD, Ciconelli RM, Carrera EF, Griffin S, Faloppa F, Dos Reis FB. Validity and reliability of the Western Ontario Rotator Cuff Index (WORC) for use in Brazil. *Clin J Sport Med*. 2008;18 :266-272. doi: 10.1097/JSM.0b013e31817282f4.

22. Lopes AD, Stadniky SP, Masiero D, Carrera EF, Ciconelli RM, Griffin S. Traducao e adaptação cultural do WORC: um questionario de qualidade de vida para alterações do manguito rotador. *Rev Bras Fisioter*. 2006;10 :309-15. doi: 10.1590/S1413-35552006000300010

23. Arcuri F, Barclay F and Nacul I. Traduccio´ n, adaptacio´ n transcultural, validacio´ n y medicio´ n de propiedades de la versio´n en espan˜ol del ı´ndice Western Ontario Rotator Cuff (WORC). *Artroscopia* 2015; 22: 56–60.

24. Zhaeentan, S., Legeby, M., Ahlström, S., Stark, A., & Salomonsson, B. A validation of the Swedish version of the WORC index in the assessment of patients treated by surgery for subacromial disease including rotator cuff syndrome. *BMC Musculoskeletal Disorders*. 2016; 17: 165. doi:10.1186/s12891-016-1014-y

25. El O, Bircan C, Gulbahar S, et al. The reliability and validity of the Turkish version of the Western Ontario Rotator Cuff Index. *Rheumatol Int*. 2006;26:1101-1108. doi:10.1007/s00296-006-0151-2

26. Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *J Clin Epidemiol*. 1993; 46:1417-32. doi: 10.1016/0895-4356(94)90091-4

27.Acquadro C, Jambon B, Ellis D, Marquis P. Language and translation issues. In: Spiker B (ed) *Quality of life pharmacoeconomics in clinical trials*. 1996. Lippincott-Raven Publishers, Philadelphia, pp 575–585

28. Longo UG, Vasta S, Maffulli N and Denaro V. Scoring systems for the functional assessment of patients with rotator cuff pathology. *Sports Med Arthrosc.* 2011;19:310–20. doi: 10.1097/JSA.0b013e31820af9b6.

29. Huber W, Hofstaetter JG, Hanslik-Schnabel B, Posch M, Wurnig C. [Translation and psychometric testing of the Western Ontario Rotator Cuff Index (WORC) for use in Germany]. Z *Orthop Ihre Grenzgeb*. 2005;143:453-60. doi:10.1055/s-2005-836677

30. Hsu, H. C., Wu, J. J., Jim, Y. F., Chang, C. Y., Lo, W. H., & Yang, D. J. Calcific tendinitis and rotator cuff tearing: a clinical and radiographic study. *Shoulder Elbow*. 1994;3:159-164. doi.org/10.1016/S1058-2746(09)80095-5

31. Jim, Y. F., Hsu, H. C., Chang, C. Y., Wu, J. J., & Chang, T. Coexistence of calcific tendinitis and rotator cuff tear: an arthrographic study. *Skeletal radiol*.1993; 2:183-185. doi: 10.1007/BF00206150

32. Garg, R. Methodology for research I. *Indian Journal of Anesthesia*. 2019;60:640–645. doi.org/10.4103

33. Reeve, B. B., Wyrwich, K. W., Wu, A. W., Velikova, G., Terwee, C. B., Snyder, C. F., ... & Lyons, J. C. ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. *Qual Life Res*. 2013;22:1889-1905. doi.org/10.1007/s11136-012-0344-y

34. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychol Methods*.1996;1:30-46. doi/10.1037/1082-989X.1.4.390

35. Anthoine, E., Moret, L., Regnault, A., Sébille, V., & Hardouin, J. B. Sample size used to validate a scale: a review of publications on newly-developed patient reported outcomes measures. *Health Qual Life Outcomes*. 2014;12:2. doi: 10.1186/s12955-014-0176-2

Gaps in current knowledge

Based on our literature review and by comparing against previously published work on the clinical measurement properties of the Short-WORC, we found the following gaps in the literature.

1) The content of the Short-WORC has never been formally assessed for content validity. It is always best practice to know the content of the PRO that a clinician is going to use, and to understand whether the items cover concepts that are important to patients.

2) The reliability (reproducibility) of the Short-WORC has not been sufficiently explored in a prospective patient population for RCDs.

3) The validity (criterion and construct) and responsiveness of the Short-WORC has not been evaluated in a prospective patient population for RCDs.

Research question

Does the Short-WORC exhibit sufficient psychometric properties when evaluated in a population of RCDs?

Objectives

The overall objective of this thesis was to evaluate the psychometric properties of the Shortened Western Ontario Rotator Cuff Index. Based on the research gaps identified in the previous sections the specific objectives are as follows:

1) To perform a comprehensive systematic review of the literature and summarize the cross-cultural adaptions and the evidence that is available on the psychometric properties of the WORC.
2) To evaluate the internal consistency and reproducibility of the Short-WORC for rotator cuff pathology.
3) To evaluate the content validity of the Short-WORC for rotator cuff pathology.
4) To evaluate the cross-sectional construct validity, longitudinal construct validity and sensitivity to change of the Short-WORC for rotator cuff pathology.

Thesis Overview

The following chapters will discuss the evaluation process of the Short-WORC. In Chapter 2, we will focus on the reliability of the Short-WORC, specifically looking at the reproducibility (reliability and agreement) parameters and the floor and ceiling effects in a prospective population. Chapter 3 will focus on the evaluation of validity (construct and longitudinal) of the Short-WORC in the population. In Chapter 4, we explore content validity of the Short-WORC through cognitive interviewing. This will fulfill the proposed objectives and overall research question. Finally, Chapter 5 will conclude with the discussion and overview of the overall work. This will further discuss the strengths, limitations, clinical and research implications, and future directions of this work.

References

1. Saunders, R., Astifidis, R., Burke, S. L., Higgins, J., & McClinton, M. A. (2015). *Hand and upper extremity rehabilitation-e-book: a practical guide*. Elsevier Health Sciences.

2. Wilson, A. (2002). *Effective management of musculoskeletal injury: A clinical ergonomics approach to prevention, treatment and rehabilitation*. Elsevier Health Sciences.

3. Hertling, D., & Kessler, R. M. (2006). *Management of common musculoskeletal disorders: physical therapy principles and methods*. Lippincott Williams & Wilkins.

4. Voss, D., Ionta, M., and Myers, B. (1985) *Proprioceptive neuromuscular facilitation*. Harpers & Row: Philadelphia.

5. Whoqol Group. (1995). The World Health Organization quality of life assessment (WHOQOL): position paper from the World Health Organization. *Social science & medicine.*41(10), 1403-1409.

6. Dewan, N., MacDermid, J. C., MacIntyre, N., & Grewal, R. (2016). Reproducibility: reliability and agreement of short version of western Ontario rotator cuff index (short-WORC) patients with rotator cuff disorders. *Journal of Hand Therapy*. 29(3), 281-291.

7. Dewan, N., MacDermid, J. C., & MacIntyre, N. (2018). Validity and Responsiveness of the Short Version of the Western Ontario Rotator Cuff Index (Short-WORC) in Patients with Rotator Cuff Repair. *Journal of Orthopedic & sports physical therapy*, 48(5), 409-418.

8. Kirkley A, Alvarez C, Griffin S. (2003). The development and evaluation of a disease-specific quality-of-life questionnaire for disorders of the rotator cuff: the Western Ontario Rotator Cuff Index. *Clin J Sport Med*. 13(2):84-92.

9. Holtby R, Razmjou H. (2005). Measurement properties of the Western Ontario rotator cuff outcome measure: a preliminary report. *J Shoulder Elbow Surg*.14(5):506- 510.

10. Furtado, R., & MacDermid, J. C. (2018). Clinimetrics: Short Western Ontario Rotator Cuff Index. *Journal of physiotherapy*. (In press)

11. Reeve BB, Wyrwich KW, Wu AW, et al. (2013) ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. *Qual Life Res*. 22(8):1889-1905.

12. Lohr KN, Aaronson NK, Alonso, J. (1996) Evaluating quality-of-life and health status instruments: Development of scientific review criteria. *Clin Ther*. 18(5):979-992.

13. Deyo RA, Diehr P, Patrick DL. (1991). Reproducibility and responsiveness of health status measures. statistics and strategies for evaluation. *Control Clin Trials*. 12(4 Suppl):142S-158S.

14. McGraw KO, Wong SP. (1996) Forming inferences about some intraclass correlation coefficients. *Psychol Methods*.1(1):30-46.

15. Portney L, Watkins M. (2000). Foundation of clinical research: Applications to practice. Upper Saddle River, NJ: Prentice-Hall.

16. Willis, G. B. (2004). *Cognitive interviewing: A tool for improving questionnaire design*. Sage Publications.

17. Thorne, S. (2016). *Interpretive description*. Routledge.

18. Magasi, S., Ryan, G., Revicki, D., Lenderking, W., Hays, R., Brod, M., . . . Cella, D. (2012). Content validity of patient-reported outcome measures: Perspectives from a PROMIS meeting. *Quality of Life Research*. 21(5), 739-746.

19. Salkind, N. J. (2010).*Encyclopedia of research design* Thousand Oaks, CA: SAGE Publications, Inc.

20. Husted JA, Cook RJ, Farewell VT, Gladman DD. (2000) Methods for assessing responsiveness: A critical review and recommendations. *J Clin Epidemiol*. 53(5):459.

21. Beaton DE, Bombardier C, Katz JN. (2001). Looking for important change/differences in studies of responsiveness. outcome measures in rheumatology. minimal clinically important difference. *J Rheumatol*. 28(2):400-405.

22. Liang MH, Larson MG, Cullen KE, Schwartz JA. (1985) Comparative measurement efficiency and sensitivity of five health status instruments for arthritis research. *Arthritis Rheum*. 28(5):542-7

23. Silverstien BA, Viikari-Juntura E, Fan ZJ.et al. (2006) Natural course of non-traumatic rotator cuff tendinitis and shoulder symptoms in a working population. *Scand J Work Environ Health*.32-99

24. White, J. J. E., Titchener, A. G., Fakis, A., Tambe, A. A., Hubbard, R. B., & Clark, D. I. (2014). An epidemiological study of rotator cuff pathology using The Health Improvement Network database. *The bone & joint journal*, *96*(3), 350-353.

25. Yamamoto, A., Takagishi, K., Osawa, T., Yanagawa, T., Nakajima, D., Shitara, H., & Kobayashi, T. (2010). Prevalence and risk factors of a rotator cuff tear in the general population. *Journal of Shoulder and Elbow Surgery*, *19*(1), 116-120.

26. Dang, A., & Davies, M. (2018). Rotator cuff disease: treatment options and considerations. *Sports Medicine and Arthroscopy Review*, *26*(3), 129-133.

27. Kuhn JE, Dunn WR, Sanders R, et al. Effectiveness of physical therapy in treating atraumatic full-thickness rotator cuff tears: a multicenter prospective cohort study. *J Shoulder Elbow Surg.* 2013;22:1371–1379.

Chapter 2: Evaluating the reproducibility of the short version of the Western Ontario Rotator Cuff Index (Short-WORC) prospectively

Abstract

Background: Recently, a shorter version of the Western Ontario Rotator Cuff Index (Short-WORC) was created to reduce patient response burden. However, it has yet to evaluated prospectively for reproducibility (reliability and agreement). Secondary outcome is to assess the floor and ceiling effects.

Methods: Patients (n=162) with rotator cuff disorders (RCDs) completed the Short-WORC at baseline and 3 months follow up (n=51). From this cohort, 47 patients underwent test-retest reliability within 2-7 days. Cronbach's alpha (a) was used to determine internal consistency and $ICC_{2,1}$ for test-retest reliability. To evaluate parameters of agreement, standard error of measurement (SEM), minimal detectable change (MDC90) and Bland Altman plots were used.

Results: Cronbach's alpha was 0.82 at baseline and 0.87 at 3 months follow up. The $ICC_{2,1}$ was 0.87. The agreement parameters were 8.4 for SEM agreement, 19.5 for MDC90 individual and 1.7 for MDC90 group. Limits of agreement fell within the range of 23.8, -22.9. No floor or ceiling effects were present.

Conclusions: The Short –WORC demonstrated strong reproducibility parameters and can be used for patients with RCDs. Wider LOA is expected for individual patient assessment, when using the Short-WORC. While it is critical to evaluate the reproducibility of an instrument, other properties such as the validity of the Short-WORC requires further evaluation.

Level of Evidence: Level II

Keywords: Rotator cuff disorders, reproducibility, agreement, reliability

A version of this work has been submitted for journal publication: Furtado R, MacDermid JC, Bryant DM, Faber K, and Athwal G. Evaluating the reproducibility of the short version of the Western Ontario Rotator Cuff Index (Short-WORC) prospectively. Journal of Shoulder and Elbow Surgeons (2019)

Introduction

Rotator cuff disorders (RCDs) are the most common cause of impairment and activity limitation, resulting in a loss of quality of life.[1] The prevalence of partial and full-thickness rotator cuff tears is greater than 60% in symptomatic patients over the age of 60.[1,2] Therefore, the primary goal of both surgery and rehabilitation is to improve the function and QoL in patients with RCDs.[1]

Recently, a shorter version of the Western Ontario Rotator Cuff Index (Short-WORC) was adapted from its original format, to evaluate QoL in patients with RCDs. Through theoretical and clinical principles supported with a factor analysis, the WORC was reduced from twenty-one items to seven items from the domains of work and lifestyle. [3,4] The Short-WORC consists of a smaller number of items that focus on activity limitations, and generates a single summary score without 5 domain scores generated by the original version of the WORC. [5] In 2012, Razmjou et al found strong psychometric properties for the Short-WORC, and suggested that it reduces response burden.[3] Shortly after, Dewan et al., found excellent reliability, validity and responsiveness when extracting scores from the full WORC.[1,6] This collection of work suggests that the Short-WORC has excellent psychometric properties when compared to the full WORC and other patient-reported outcomes (PROs).[5] However, there are no studies that have prospectively evaluated the reproducibility (reliability and agreement) of the Short-WORC . [1,3,6]

Reproducibility measures the extent to which similar results are obtained from repeated assessments. Furthermore, reproducibility is a broad term that incorporates the parameters of both reliability and agreement.[7-9] Reliability focuses on the degree to which test scores are consistent, dependable, repeatable, and to a degree, free of measurement error. Reliability can be further investigated through internal consistency (cross-sectional reliability) and test-retest reliability (longitudinal reliability).[7-9] Additionally, the property of agreement focuses on measurement error and evaluates the proximity of scores derived from repeated measurements. Agreement is investigated through absolute reliability coefficients (standard error of the measurement, minimal detectable change) and Bland Altman (BA) plots. [9,10]

PROs must demonstrate both reliability for discriminative applications, and agreement to discern real change from error.[7,10] Therefore, it is critical to examine both reliability and agreement

in outcome measures.[1] Thus, the purpose of this study was to evaluate reproducibility (internal consistency, test-retest reliability and agreement) of the Short-WORC in a prospective patient population with RCDs.

Methods

Study Design

The reproducibility (internal consistency, test-retest reliability and agreement) of the Short-WORC was assessed through a prospective cohort of patients undergoing treatment at the Roth McFarlane Hand and Upper Limb Centre, at St. Joseph's Health Care London, London, Canada. Ethics approval was granted by the Western University Research Ethics Board.

Participants

Prospective data collection of patients over the age of 18 years and diagnosed at the Hand and Upper Limb Centre with a rotator cuff disorder were eligible for the study. Patients who had upper extremity fractures, adhesive capsulitis, shoulder instability, infection, tumors, labral, cartilage, and ligamentous tears were excluded from the study. Patients (n=162) with completed individual items scored on the Short-WORC at baseline and n= 51 at 3 months follow up were included. Out of the 162 patients at baseline, 47 stable participants were retested within 2-7 days for test-retest reliability.

We expect to obtain test-retest reliability and internal consistency (ICC) of 0.90 as shown in previous studies[1,29]. Therefore, the sample size required to determine whether the reliability of the Short-WORC exceeds 0.95 CI around a power of 0.80.[30]

Outcome Measures

The 7 item Short-WORC was originally shortened and validated by Razmjou et al, from the domains of work and lifestyle.[3] The Short-WORC total score can range from 0 (best possible score) to 700 (worst possible score). The percentage score is obtained from the sum of the raw item scores, divided by 700 and multiplied by 100. This generates a score between 0 (poor QoL) to 100 (high QoL) percent. The Short-WORC cannot be scored if items are missing. [3,5]

Statistical Analysis

Data was assessed for completeness, percentage of missing data, presence of outliers and floor and ceiling effects. The data set was tested for normality, however, showed to be non-normal. However, according to the central limit theorem, the distribution of means from any non-normal distribution can still be considered approximately normal as long as samples (n) are larger than 30 participants. [27] Therefore, we used parametric statistics for our analysis as our sample is greater than 30 participants.

SPSS version 24 (IBM, SPSS Inc., Chicago, IL) software was used for data analysis and a p-value of < 0.05 was considered statistically significant.

Floor and ceiling effects (F/C effects)

The floor and ceiling effects (F/C effects) were calculated by the percentage of patients whose total score fell between 0 and 10 (minimal scores) and 90 to 100 (maximal scores). As suggested by McHorney and Ware, F/C effects are defined by using a cut-off of 15%.[11] Therefore, F/C effects were considered to exist if >15% of participants scored minimal or maximal total scores.

Reliability

Internal consistency is defined as the extent to which items in the questionnaire are correlated with each other, when assessed at one point in time.[12] Cronbach's alpha was calculated with a 95% CI to assess internal consistency at baseline and at three months of follow up. An alpha of 0.70-0.90 was deemed as having excellent internal consistency.

Test-retest reliability (longitudinal reliability) measures the extent to which consistent results are obtained at test and retest occasions in stable subjects.[13] A value of 0.70- 0.80 is deemed appropriate for comparison in research, and over 0.90 for clinical interpretation. [12,21] Test-retest scores were analyzed using a 2-way mixed model with absolute agreement to produce an intra-class correlation ($ICC_{2,1}$), with a 95% CI for a single measure. [14,15] An ICC of 0.80 was considered as a minimum standard for good reliability in this study.[1,14]

Statistical hypothesis

We expect that the Short-WORC will demonstrate excellent internal consistency (Cronbach's alpha) and test-retest reliability ($ICC_{2,1}$) $\geq$ 0.80 and $\geq$ 0.90, respectively.

Agreement parameters

Absolute reliability was assessed by calculating a standard error of measurement (SEM) and minimal detectable change (MDC_{90}) statistic. The SEM was calculated using the following equation[16]:

$$SEM_{agreement} = \text{Standard Deviation}_{pooled} \times \sqrt{1\text{-}ICC_{2,1\ agreement}};$$

$$\text{where Standard Deviation}_{pooled}\ (SD_{pooled}) = SD_{test} + SD_{retest}\ /2$$

Assuming that our data verifies the two required assumptions for estimation of MDC_{90} (i.e., no systematic bias and normally distributed data), we used SEM to calculate the MDC_{90}, using the following equation[1,9]:

$MDC_{90} = 1.64\ X\ SEM_{absolute\ agreement} \times \sqrt{2}$.

SEM provides the estimate of measurement error in the same units as the original measurement and MDC_{90} is the minimum amount of change which is required to be 90% confident that a change has occurred over a period of time without measurement error.[9] The 95% CI for MDC_{90} was calculated by[1,9]:

95% CI for $MDC_{90} = d \pm MDC_{90,}$ as "d" is the mean difference.

To calculate the real change over time between groups of patients, we calculated for MDC_{group}, using the formula [1,16,17]:

$MDC_{group} = MDC_{90} / \sqrt{n} \times 1.64$, where n = sample size of group.

Smaller SEM and MDC values indicate smaller measurement error.[9]

Bland Altman Plot (BA plot)

The BA plots were used for plotting the difference between scores at time one and two of the test-retest period against their mean score for the two points with 95% limits of agreement (LOA). The BA plots produces an image of the results that can be used to evaluate systematic variability (bias), the present of outliers, and homoscedasticity. [18-20]

Results

The demographics of the study population are presented in Table 1.

Reliability

There were no floor or ceiling effects. Internal consistency (Cronbach's alpha (CI 95%)) was excellent at the baseline assessment n=162 (0.82) and three months postoperatively n=51 (0.87).

Test-retest reliability

Test-retest reliability was excellent ($ICC_{2,1}$=0.87).

Agreement parameters

Reported values for the $SEM_{agreement}$ (8.4), MDC_{90} (19.5) and $MDC_{90group}$ (1.7) are reported in Table 3.

Bland Altman plot (BA plot)

The 95% LOA for test-retest scores are presented in Table 2. Visual inspection shows the random scatter of most points to be within the 95% LOA and represents negligible systematic bias between scores for the Short-WORC (see Figure 1).

Discussion

This study demonstrated excellent reliability and agreement properties for the Short-WORC when administered to a group of patients with RCDs. Our findings provide strong evidence to support the finding of previous studies that assessed these properties retrospectively.[1,3] Together,

this collection of studies suggests that the Short-WORC is sufficiently reproducible such that clinicians can have confidence in the stability of patient scores. [1,3,6] when making decisions about patient quality of life and changes in quality of life

In this study, we did not observe floor or ceiling effects, which is also consistent with previously published work[1,3,6] suggesting, that the Short-WORC is well suited to detect both improvement and worsening in the RCD population. The internal consistency (0.82) was both acceptable and similar to that reported by Razmjou et al. (0.89), and Dewan et al. (0.84) at baseline [1,3] and comparable to the Cronbach's alpha (0.85-0.92) of the original WORC depending on the translation.[22-24] Because it is suggested that values exceeding 0.90 indicate redundancy, the Short-WORC may be more efficient than the WORC.[1,3]

While an ICC of 0.90 or greater can be difficult to obtain, previous literature considers a measure reliable if the point estimate exceeds 0.75.[1,12,21] The ICCs found in this study were similar to those of previously published work and the WORC.[3,22-24] Since the $ICC_{2,1}$ value (0.87) exceeds the benchmark of 0.75, our study provides strong evidence that the Short-WORC has excellent reliability across multiple contexts. Based on our narrow CI, we can be confident that our estimate is precise and exceeds minimum expectations.

The $SEM_{agreement}$ of 8.4 for the Short-WORC reported in our study, indicated that there was a 68% chance ($1 \pm SEM$) that true scores of Short-WORC for an individual assessed at a single point in time lies within 8.4 points of the measured score. We have used the $ICC_{2,1\ absolute\ agreement}$ to calculate SEM instead of Cronbach's alpha, and did not choose to use Cronbach's alpha to estimate SEM. Instead, we used $SEM_{agreement}$ to compute MDC,[1,12,25] as it expresses the measurement error through the systematic difference between test and retest scores, which are otherwise ignored with $SEM_{consistency}$. [25]

The MDC_{90} of the Short-WORC implies that if the individual's score on the Short-WORC has changed by at least 19.5 points, the clinician can be confident that true change (over and above questionnaire error) has occurred. In comparison to the WORC (17.8)[1], we see that the MDC values are higher for the Short-WORC (19.5). This could be a result of fewer items in the Short-WORC, therefore, producing greater variability.

The low value of the MDC_{group} indicates that the Short-WORC is an excellent measure of change within a group of patients. When comparing both the MDC_{90} values (individual vs group), the Short-WORC is better at measuring change for a group of patients than the individual.[26] As shown in the literature, a smaller value of the $MDC_{90group}$ than the $MDC_{90individual}$ aligns with agreement parameters reported for other patient reported outcome measures. This is an expected finding since the $MDC_{90group}$'s formula is dependent on the square root of the sample size, unlike the $MDC_{90individual}$ which is dependent on the square root of 2 and the error band around the mean difference of two measurements. This is further evident as the group effect will always average out any differences that would be normally highlighted in the individual effect. Therefore, the variability will always be higher for the $MDC_{90invidiual}$ compared to the $MDC_{90group}$. However, measuring both group and individual is important to assure that the measure is reliable when assessing an individual patient over an interval of test-retest, and over a period of time between groups of patients after an intervention.[1,7]

The LOA on BA plots are known to represent the interval within which repeated measures would be expected to fall 95% of the time. The wide 95% LOA (23.8, -22.9) reported in our study reflects large within individual variability and hence limited usefulness of measures for individual comparisons. We used the retest assessment of 2-7 days as a stable time-period for patients, as it is long enough to prevent recall bias but short enough to expect that no clinical change has occurred since RCDs are a chronic condition. This interval was sufficient according to other previous literature, but can allow some potential for circumstances to de-stabilize the patient's condition.[1] Our assumption of considering one week as the time interval was supported by the results of the BA plot, indicating a stable time frame.

The negligible mean difference and acceptable agreement of the Short-WORC reported in the present study, suggests that the Short-WORC can replace the 21-item WORC for both clinical and research application. However, while there are high values of the LOA (23.8, -22.9), they are similar to those of the WORC (20.1, -22.7), and our previously published work (22.3, -26.5)[1]. The agreement parameters are also in accordance with our previous published work. [1,3,6]

Overall, our findings are consistent with values obtained when the Short- WORC was extracted from its full parent version. Lower internal consistency and wider variations between test-

retest scores reliability can be expected when using abbreviated questionnaires.[28] The goal of shortening the questionnaire is to reduce patient and/or administrative burden while retaining the conceptual linkage to the intended construct and sufficient psychometric equivalence. Although we did not directly measure the time, we assumed that patients required less time to complete the 7-items of the Short-WORC compared to the 21-items of the original WORC. In this study, only certain psychometric properties of the Short-WORC were assessed. Therefore, future studies should evaluate comprehension and construct clarity of the Short-WORC through qualitative studies, and longitudinal studies of responsiveness. Although our previous work supports the responsiveness of the Short-WORC, it was conducted using data collected from the original version of the WORC. Therefore, it is important to understand whether the equivalence between the extracted and isolated versions of the Short-WORC are consistent. As well, all of the studies to date have been conducted at specialty shoulder surgery clinics, therefore, assessment in populations of different contexts or that include a broader spectrum of RCD would clarify whether these measurement properties exist in multiple contexts of the disorder.

Conclusion

The Short-WORC has an absence of ceiling and floor effects, acceptable internal consistency, excellent reliability for group comparisons; and suitable, but imperfect confidence in the test-retest reliability of scores at the level of the individual patient with RCD. While reproducibility data are essential, data to evaluate the validity and responsiveness of the Short-WORC are still required.

References

1. Dewan, N., MacDermid, J. C., MacIntyre, N., & Grewal, R. Reproducibility: Reliability and agreement of short version of Western Ontario Rotator Cuff Index (Short-WORC) in patients with rotator cuff disorders. *Journal of Hand Therapy*, 29 (2016), PP. 281–291

2. MacDermid JC, Ramos J, Drosdowech D, Faber K, Patterson S. The impact of rotator cuff pathology on isometric and isokinetic strength, function, and quality of life. J Shoulder Elbow Surg. 2004 Nov-Dec;13(6):593-598.

3. Holtby R, Razmjou H. Measurement properties of the Western Ontario rotator cuff outcome measure: a preliminary report. J Shoulder Elbow Surg. 2005 Sep- Oct;14(5):506- 510.

4. Kirkley A, Alvarez C, Griffin S. The development and evaluation of a disease- specific quality-of-life questionnaire for disorders of the rotator cuff: theWestern Ontario Rotator Cuff Index. Clin J Sport Med. 2003 Mar;13(2):84-92.

5. Furtado, R., & MacDermid, J. C. (2018). Clinimetrics: Short Western Ontario Rotator Cuff Index. *Journal of physiotherapy*. (In press)

6. MacDermid, J. C., Khadilkar, L., Birmingham, T. B., & Athwal, G. S. (2015). Validity of the QuickDASH in patients with shoulder-related disorders undergoing surgery. *Journal of orthopaedic & sports physical therapy*, *45*(1), 25-36.

7. de Vet HC, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. J Clin Epidemiol. 2006 Oct;59(10):1033-1039.

8. Streiner DL, Norman GR. Health Measurement Scales. A Practical Guide to Their Development and Use. 4th ed. *Oxford: Oxford University Press*; 2008.

9. Beckerman H, Roebroeck ME, Lankhorst GJ, Becher JG, Bezemer PD, Verbeek AL. Smallest real difference, a link between reproducibility and responsiveness. *Qual Life Res*. 2001;10(7):571-578.

10. Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. *J Chronic Dis*. 1987;40(2):171-178.

11. McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Qual Life Res*. 1995 Aug;4(4): 293-307.

12. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007 Jan;60(1):34-42.

13. Portney LG, Watkins MP. Foundations of Clinical Research: Applications to Practice. Upper Saddle River, NJ: Prentice Hall Health; 2009.

14. Schuck P. Assessing reproducibility for interval data in health-related quality of life questionnaires: which coefficient should be used? *Qual Life Res*. 2004 Apr;13(3):571-586

15. McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Qual Life Res*. 1995 Aug;4(4): 293-307.

16. Chen HM, Hsieh CL, Sing Kai L, Liaw LJ, Chen SM, Lin JH. The test-retest reli- ability of 2 mobility performance tests in patients with chronic stroke. *Neurorehabil Neural Repair*. 2007 Jul-Aug;21(4):347-352.

17. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979 Mar;86(2):420-428.

18. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res*. 1999 Jun;8(2):135-160.

19. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986 Feb 8;1(8476):307-310.

20. Brennan P, Silman A. Statistical methods for assessing observer variability in clinical measures. *BMJ*. 1992 Jun 6;304(6840):1491-1494.

21. Kottner J, Audige L, Brorson S, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *J Clin Epidemiol*. 2011 Jan;64(1): 96-106.

22. Kawabata M, Miyata T, Nakai D, Sato M, Tatsuki H, Kashiwazaki Y, Saito H. Reproducibility and validity of the Japanese version of the Western Ontario Rotator Cuff Index. *J Orthop Sci*. 2013;18:705–11.

23. El O, Bircan C, Gulbahar S, et al. The reliability and validity of the Turkish version of the Western Ontario Rotator Cuff Index. *Rheumatol Int*. 2006 Oct;26(12):1101-1108.

24. Diniz Lopes A, Ciconelli RM, Carrera EF, Griffin S, Faloppa F, Baldy dos Reis F. Comparison of the responsiveness of the Brazilian version of the Western Ontario Rotator Cuff Index (WORC) with DASH, UCLA and SF-36 in patients with rotator cuff disorders. *Clin Exp Rheumatol*. 2009 Sep-Oct;27(5):758-764.

25. Harvill LM. Standard error of measurement. *Educ Meas Issues Pract*. 1991;10: 33-41.

26. Streiner DL, Norman GR. Health Measurement Scales. A Practical Guide to Their Development and Use. 4th ed. *Oxford: Oxford University Press*; 2008.

27. Norman GR, Streiner DL. The Normal Distribution. *Biostatistics: The Bare Essentials.* Hamilton, Ont.: B.C. Decker; 2008:32.

28. Beaton, D. E., Katz, J. N., Fossel, A. H., Wright, J. G., Tarasuk, V., & Bombardier, C. (2001). Measuring the wole or the parts?: Validity, reliability, and responsiveness of the disabilities of the arm, shoulder and hand outcome measure in different regions of the upper extremity. *Journal of Hand Therapy*, *14*(2), 128-142.

29. Fowler JJP., Chevannes M. Practical statistics for nursing and health care. West Sussex, UK: *John Wiley & Sons*; 2002

30. Bonett, D. G. (2002). Sample size requirements for estimating intraclass correlations with desired precision. *Statistics in medicine*, *21*(9), 1331-1335.

Table 1. Patient baseline characteristics (n= 162)

Variable	N/percentage
Age in years (mean SD)	(61.2 ± 16.3)162/100
Sex	
Males	72 / 44.4
Females	90 / 55.5
Affected shoulder	
Left	54 / 33.3
Right	94 / 58.1
Bilateral	14 / 8.6
Occupational	
Employed	80/49.4
LOA	12/7.4
Unemployed	10/6.2
Retired	60/37.1

Table 2. Longitudinal reliability of the Short-WORC

Test-retest reliability					
Test mean (SD)	**Retest mean (SD)**	**d (SD)**	**95% CI**	**95% LOA**	**ICC (95%CI)**
45.6 (23.9)	45.1 (23.1)	0.5 (11.9)	-2.8, 3.4	23.8, -22.9	0.87 (0.79, 0.92)

Table 3. Reproducibility: Agreement parameters of the Short-WORC

SEM $_{\textbf{agreement}}$	**MDC**$_{\textbf{90individual}}$	**MDC**$_{\textbf{90group}}$
8.4	19.5	1.7

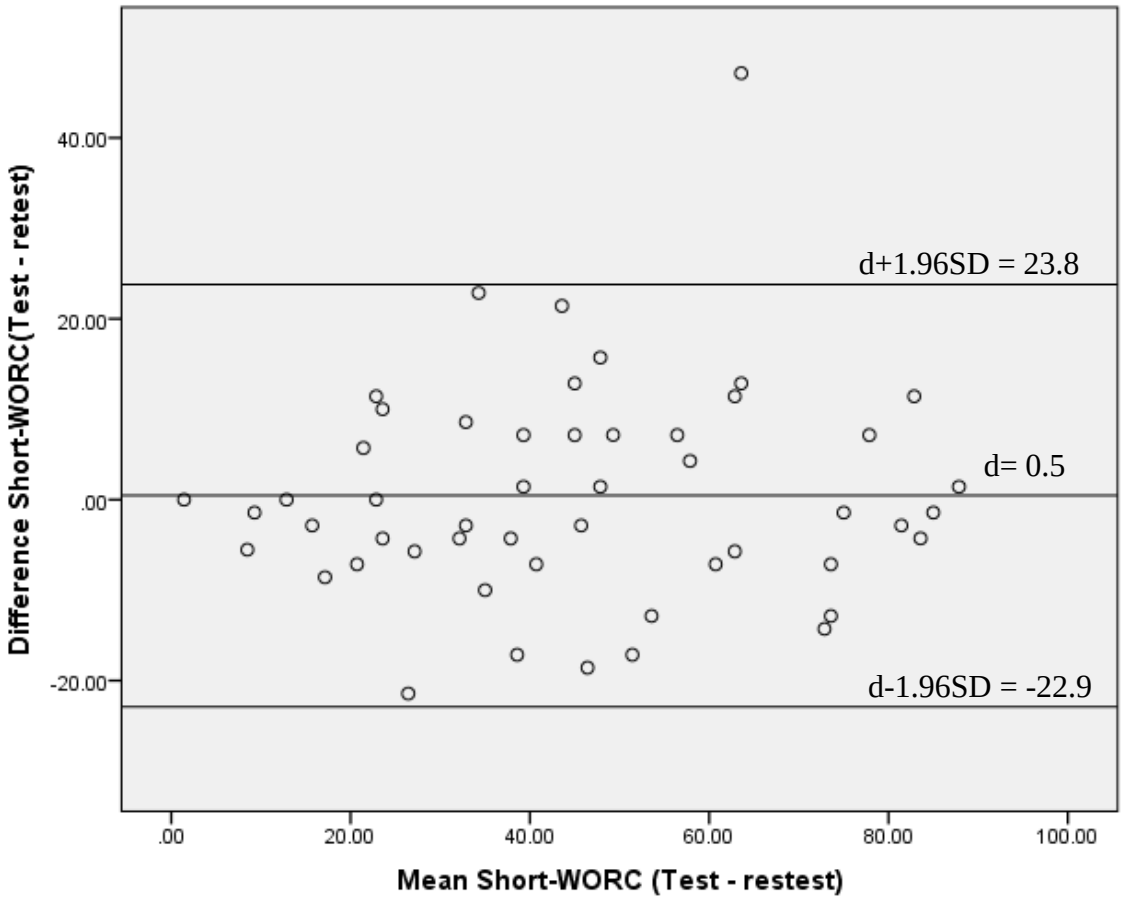

Figure 1. Bland Altman Limits of Agreement (LOA) Plot between test and retest of the Short-WORC. n = 51 The central line represents the mean of the individual differences (d) and two lines to represent the 95% LOA. Differences lie between d $\pm$ 1.96SD of mean difference.

Chapter 3: Evaluating the validity and the sensitivity to change of the Short version of the Western Ontario Rotator Cuff Index (Short-WORC) prospectively

Abstract

Background: Recently, a shorter version of the Western Ontario Rotator Cuff Index (Short-WORC) was created to reduce patient response burden. However, it has yet to be prospectively evaluated.

Purpose: To evaluate the validity and the sensitivity to change of the Short-WORC with comparator measures (Simple Shoulder Test; (SST), American Shoulder and Elbow surgeon shoulder score; (ASES), EQ-5D and Short-form McGill Pain Questionnaire; (SF-MPQ-2) in patients undergoing rotator cuff repair.

Methods: Patients (n=162) with rotator cuff disorders (RCDs) completed the Short-WORC at baseline and 3 months (n=72) and 6 months (n=30) follow up appointments. The construct validity (Pearson correlations), sensitivity to change (effect size), standardized response mean (SRM) and relative efficiency (RE)) of the Short-WORC were measured.

Results: The Short-WORC was able to discriminate sex difference and between health status at both the cross-sectional level and at the longitudinal level. The Short-WORC had a moderate correlation with other comparator measures, [SST (r= 0.55), and SF-MPQ-2 (r= 0.50-0.55)] but also some weaker correlations with the ASES (r =0.3-0.5) and EQ-5D-5L (r= 0.3-0.55). The Short-WORC demonstrated the ability to measure change at 0-3 months (ES=0.1, SRM=0.1), and at 0-6 months (ES=0.6, SRM=0.6), but the reported scores were lower than the other comparator measures.

Conclusions: The Short-WORC when measured prospectively is able to discriminate sex and health status subgroups, and provides some evidence that it can detect change over time. However, the Short-WORC did not report similar values of sensitivity to change, when compared to other comparator measures. Therefore, future studies should focus on measuring the Short-WORC in an all surgical patient population to confirm that it contains the same validity properties as the WORC.

Keywords: Rotator cuff disorders, patient reported outcomes, quality of life, sensitivity to change, validity

Introduction

Rotator cuff disorders (RCDs) are the most common cause of a reduced quality of life (QoL) in upper limb extremities, as they cause impairment and activity limitation in patients over the age of 60.[1] Therefore, patients are directed towards nonsurgical or surgical treatments to improve their function and QoL.[1,2] Clinicians and researchers will often use disease-specific patient reported outcomes (PROs) to assess patients' recovery from treatment intervention. The Western Ontario Rotator Cuff Index (WORC), is one of the most validated disease-specific PROs for patients with RCDs.[3]

Recently, a shorter version of the Western Ontario Rotator Cuff Index (Short-WORC) was adapted from its parent, to evaluate QoL in patients with RCDs.[4,5] The Short-WORC evaluates the activity limitations that arise from RCDs, and focuses on seven items from the domains of work and lifestyle. As our previous work shows[2,6], the Short-WORC demonstrates strong reliability, validity and responsiveness properties when the scores were extracted from the full WORC. Furthermore, the findings suggest that the psychometric properties of the Short-WORC are comparable to the full WORC and other shoulder patient-reported outcomes (PROs).[2,6] However, the previously published studies on the Short-WORC have yet to evaluate validity and sensitivity to change prospectively or when administering the actual Short-WORC instead of the parent version.

PROs must demonstrate strong validity when making decisions regarding clinical change over time.[6] Therefore, it is critical to examine all facets of validity to verify that the shortened version retains the construct of the original. Thus, the purpose of this study was to evaluate the validity (cross-sectional and longitudinal) of the Short-WORC after 6 months in a prospective patient population with RCDs.

Methods

Study Design

The validity of the Short-WORC was assessed through a prospective cohort of patients undergoing treatment at the Roth|McFarlane Hand and Upper Limb Centre, at St. Joseph's Health Care London, London, Canada. Ethics approval was granted by the Western University Research Ethics Board.

Participants

Patients over the age of 18 years and diagnosed at the Roth | McFarlane Hand and Upper Limb Centre, with a rotator cuff disorder were eligible for the study. Patients who had upper extremity fractures, adhesive capsulitis, shoulder instability, infection, tumors, labral, cartilage, and ligamentous tears were excluded from the study. Patients (n=162) with completed individual items scored on the Short-WORC at baseline, 3 months (n= 72) and 6 months (n=30) follow up were included.

Outcome Measures

The 7 item Short-WORC, containing items from the domains of work and lifestyle, was originally shortened and validated by Razmjou et al.[4] The Short-WORC total score can range from 0 (best possible score) to 700 (worst possible score). The percentage score is obtained from the sum of the raw item scores, divided by 700 and multiplied by 100. This generates a score between 0 (poor QoL) to 100 (high QoL) percent, and cannot be scored if items are missing.[4,5]

The American Shoulder and Elbow Surgeons scale (ASES) is a joint-specific scale that measures functional limitations and pain in the shoulder. Focusing on pain, instability and activities of daily living, the ASES is scored through both "yes/no" questions and a 0-3 numeric point scale in order to generate a score out of a total of 100 points. [7,8]

The Short-form McGill Pain Questionnaire (SF-MPQ-2) is a valid PRO that focuses on capturing the neuropathic and non-neuropathic pain conditions for a spectrum of disorders. It is comprised of 22 items, and is scored on a 0-10-point numeric scale to generate a total score out of 10.[9]

The Simple Shoulder test (SST) is a reliable and valid shoulder specific outcome measure that evaluates the change of the shoulder over time. The SST measures the functional limitations of the affected shoulder in patients with shoulder dysfunction, and contains 12 "yes/no" questions, to generate a score of shoulder limitation.[10,11]

The EuroQol 5 dimensions (EQ-5D-5L) is a valid and reliable measurement for generic health status in a variety of patient populations. Divided into 5 dimensions, patients rate their score on a 4-point Likert scale to generate a total score out of 1 to represent the best or worst health possible. [12]

The global rate of change (GRoC) scale is a valid and reliable measure that is commonly used in musculoskeletal research for determining the effect of an intervention or course of a clinical condition. The GRoC is comprised of a 7-point scale to determine whether a patient is worse, better or unchanged in health status.[13]

Statistical Analysis

Data was assessed for completeness, percentage of missing data and presence of outliers. SPSS version 24 (IBM, SPSS Inc., Chicago, IL) software was used for data analysis and a p-value of < 0.05 was considered statistically significant. Floor and ceiling effects have shown to be absent and have already been calculated in our previous work.[14]

Validity

Construct (Discriminative) validity: We assessed construct validity by evaluating the extent to which the Short-WORC can demonstrate an association between disability, function and QoL, and distinguish between known groups that are different based on theory or research. [15]

1. *Known-group validity*: As suggested in previous literature, patients who score lower on the GRC [13] demonstrate a poorer recovery. Furthermore, it is shown that females have poorer outcomes after rotator cuff repair.[2,6,16] Therefore, we evaluated the cross-sectional and longitudinal known-group validity using the variables GRoC status and sex.
    - Cross-sectional known-group validity: Independent t-tests were used to evaluate whether Short-WORC scores differ based on the GRoC status (changed vs unchanged) and sex (female vs male).

- Longitudinal known-group validity: Independent t-tests were used to evaluate longitudinal effects of the Short-WORC across the time points of baseline to 3 months follow-up and then baseline to 6 months follow up for both known groups of sex and GRoC status.

2. *Convergent construct validity*[17,18]*:* was evaluated by correlating the Short-WORC scores with measures that assess similar constructs.
    - Cross-sectional construct validity: Pearson correlation coefficients (r) with 95% CI were calculated to evaluate the relationships of the Short-WORC with other shoulder outcome measures (ASES, SST, MC-GILL-SF and EQ5D) at baseline and follow up. Correlation coefficients of very weak (0.00-0.19), weak (0.20-0.39), moderate (0.40-0.69), strong (0.70-0.89) and very strong (0.90-1.00) were defined.

We tested the following a priori hypothesis:

The Short-WORC will have a moderate correlation (0.4-0.6) with other patient reported outcomes at the baseline visit and at 3 and 6 months follow up visits.

<u>Longitudinal Validity</u>

1. Paired t-tests were used to test changes in Short-WORC between baseline and 3 and 6-month follow up appointments.[19,22] The paired t-test helped determine the relative efficiency (RE) [20] of the Short-WORC to the other shoulder PROs. RE was calculated by:

    $RE_{Short\text{-}worc} / X_1 = (t_{Short\text{-}WORC} / t_{X1})^2$

    Where, X1 = the Shoulder PRO

    t = mean difference / (SD of mean difference /$\sqrt{n}$)

    An RE > 1 indicates that the Short-WORC was a more efficient tool for measuring change in comparison to the other comparator PRO. The RE < 1 indicates less efficiency.
2. Effect size (ES I) [21,22] also known as standardized effect size, is the ratio of mean change scores (δ_x = x2- x1) to the standard deviation of the baseline scores ($SD_{baseline}$) when δ_x is mean change and x1 and x2 represent mean scores assessed at baseline and follow up

assessments respectively. A higher level of variability at baseline in relation to mean score changes will result in a smaller effect size. A trivial effect size is less than 0.2, a small effect size is between 0.2 – 0.5, a moderate effect size is 0.5-0.8, and a large effect size is $\geq$ 0.8. [23]

$$ES = \frac{\text{(mean follow-up score) - (mean baseline score)}}{\text{SD of baseline scores}}$$

3. Effect size II (ES II) [22] also known as Standardized response mean (SRM), is defined as the ratio of mean change scores ($\delta_x = x2 - x1$) to the standard deviation reflecting the variability of change scores (SD δ_x). A higher level of variability in change scores in relation to the mean change will have a smaller SRM value. ES II provides an estimate of change in the measure, standardized relative to the between patient variability in change scores. A trivial effect size is less than 0.2, a small effect size is between 0.2 – 0.5, a moderate effect size is 0.5-0.8, and a large effects size is $\geq$ 0.8. [23]

$$SRM = \frac{\text{(mean follow-up score) - (mean baseline score)}}{\text{SD of change scores}}$$

Therefore, we expect smaller effect sizes between 0-3 months since recovery from surgical repair would be incomplete, compared to at 6 months, where recovery is more complete.

Results

Table 1. Patient baseline characteristics (n=162)

Variable	N/percentage
Age in years (mean SD)	61.2 (± 16.3)
Sex	
Males	79 / 48.8
Affected shoulder	
Left	54 / 33.3
Right	94 / 58.1
Bilateral	14 / 8.6
Occupational	
Employed	80/49.4
LOA	12/7.4
Unemployed	10/6.2
Retired	60/37.1

Table 2: Cross-sectional known group validity of group 1: Sex and group 2: Health status (GRoC).

Time point	**Female (Mean(SD))**	**Male (Mean(SD))**	**Mean difference**	**p-value**
Baseline (females=84 males= 80)	30.3 (21)	43.2 (22)	12.9	0.001
Month 6 (females=17 males =14)	33.1 (25)	46.3 (28)	13.7	0.05
Time point	**GRoC change (Mean(SD))**	**GRoC no change (Mean(SD))**	**Mean difference**	**p-value**
Baseline (change= 70 no change=76)	42.8 (25)	33.3 (19)	9.5	0.03
Month 6 (change=16 no change=16)	46.3 (28)	31.7 (23)	14.6	0.12

Table 3: Longitudinal known group validity of group 1: Sex and group 2: Health status (GRoC)

Time point	**Female (Mean(SD))**	**Male (Mean(SD))**	**Mean difference**	**p-value**
0-3 (females=84 males= 80)	4.1 (1.6)	4.8 (0.1)	0.7	0.05
0-6 (females=17 males =14)	-3 (4)	-3.7 (6)	6.7	0.03
Time point	**GRoC change (Mean(SD))**	**GRoC no change (Mean(SD))**	**Mean difference**	**p-value**
0-3 (change= 70 no change=76)	6.1 (4.5)	5.5 (0.2)	0.65	0.001
0-6 (change=16 no change=16)	-3.7 (3)	1.7 (1)	-5.4	0.01

Table 4: Cross-sectional convergent construct validity. Pearson r±95% CI

Time point	**Measure**	**ASES**	**SST**	**SF-MPQ-2**	**EQ-5D-5L**
Baseline	Short-WORC	0.34	0.55*	0.49*	0.33
Month 3	Short-WORC	0.53*	0.50*	0.53*	0.45*
Month 6	Short-WORC	0.30	0.56*	0.55*	0.55*

* moderate correlation between the Short-WORC, a priori hypothesis was supported

Table 5: Longitudinal validity: mean change, SRM, ES

Change Interval	**Measure**	**t_0**	**$Mean_{baseline}$ ($SD_{baseline}$)**	**$Mean_{followup}$ ($SD_{follow-up}$)**	**Mean change (Δ)**	**SDΔ**	**SRM (Δ/SDΔ)**	**ES (Δ/$SD_{baseline}$)**
0-3 Month (N=72)	Short-WORC	0.9	31.6(22)	33.2(22)	2	20	0.1	0.1
	ASES	2.7	19.7(12.8)	25.5(15)	5.9	18	0.3	0.5
	SST	1.2	17(23)	20.5(22)	3.5	26	0.1	0.1
	SF-MPQ-2	1.9	2.7(2)	3.8(3)	1.1	4	0.5	2
	EQ-5D-5L	1	0.73(0.2)	0.71(0.1)	0.02	0.2	0.05	0.05
0-6 Month (N=30)	Short-WORC	3.4	29.1(21)	41.6(27)	12.5	20	0.6	0.6
	ASES	17	17.4(15)	31.2(19)	13	4	3.1	0.9
	SST	10	20(25)	29.3(26)	9.6	5	1.8	0.4
	SF-MPQ-2	18	2.2(1.8)	4.2(3)	2	0.6	3.2	1.1
	EQ-5D-5L	13	0.73(0.2)	0.78(1)	0.04	0.02	2	0.1

Table 6: Relative Efficacy of the Short-WORC versus other shoulder PRO to detect change over time

Change Interval	**RE$_{\text{Short-WORC/ASES}}$**	**RE$_{\text{Short-WORC/SST}}$**	**RE$_{\text{Short-WORC/SF-MPQ-2}}$**	**RE$_{\text{Short-WORC/EQ-5D-5L}}$**
Month 0-3	0.3	0.7	0.5	0.9
Month 0-6	0.2	0.4	0.2	0.3

The baseline characteristics of the study participants for each variable are available in Table 1. Since not all participants completed all measures at all time points, the sample size for each evaluation is reported within each analysis.

Cross-sectional known group validity

In table 2, the Short-WORC is able to discriminate between females at baseline [30.3(21)] and 6 months [33.1(25)] and males at baseline [43.2(22)] and 6 months [46.3(28)].

At baseline and 6 months, females reported lower scores than males which supports our hypothesis.

The Short-WORC was also able to discriminate between GRoC health status of patients who have changed at baseline [42.8(25)] and 6 months [46.3(28)] and who have not changed at both baseline [33.3(19)] and 6 months [31.7(23)]. At both time points, patients who reported change had lower scores on the Short-WORC which supported our hypothesis.

Longitudinal group validity

In table 3, the Short-WORC scores are summarized across the changes from baseline to 3 months and from baseline to 6 months. We see that in the sex group, females [4.1(1.6); -3(4)] reported a lower change in all time points than males [4.8(0.1); -3.7(6)]. Furthermore, we see that in the GRoC health status groups, patients who reported no change [6.2(4.5); -3.7(3)] demonstrated lower change in all time points than patients who reported change [5.5(0.2); 1.7(1)].

Cross-sectional convergent construct validity

Pearson's correlation coefficient for the Short-WORC with respect to the ASES, SST, SF-MPQ-2 and EQ-5D-5L is reported in Table 4. Results show that the Short-WORC was moderately correlated with the ASES at 3 months (r=0.53) which supports our hypothesis, but fails to support at baseline (0.34) or 6 months (r=0.30). The Short-WORC was moderately correlated with the SST at baseline (r=0.55), 3 months (r=0.50) and 6 months (r=0.56) which supports our hypothesis. The Short-WORC was moderately correlated with the SF-MPQ-2 at baseline (r=0.50), 3 months (r=0.53) and 6 months (r=0.55) which supports our hypothesis.

Furthermore, the Short-WORC was moderately correlated at 3 months (r=0.45) and 6 months (r=0.55) supporting our hypothesis, but was weakly correlated at baseline (r=0.33).

Longitudinal Validity (Sensitivity to Change)

Table 5 demonstrates the sensitivity to change for all PROs assessed at baseline and follow-up. All PROs were able to detect change over time when assessed at all time points. Furthermore, the ES and SRM provide evidence that the Short-WORC can measure change. As predicted, the Short-WORC demonstrated evidence of longitudinal validity at 0-3 months (ES=0.1, SRM=0.1), and a

larger change at 0-6 months (ES=0.6, SRM=0.6). The ASES (ES=0.9, SRM=3.1), SST (ES=0.4, SRM=1.8), EQ-5D-5L (ES=0.1, SRM=2) and SF-MPQ-2 (ES=1.1, SRM=3.2) were better able to detect change from 0-6 months. Table 6 demonstrates the RE of the Short-WORC to evaluate the change over time in comparison to the other comparator PROs. When compared to the SST ($RE_{3months} = 0.7$, $RE_{6months} = 0.4$) and EQ-5D-5L ($RE_{3months} = 0.9$, $RE_{6months} = 0.3$), the Short-WORC is more efficient in measuring change as RE > 1 at 3 months. With regards to the ASES ($RE_{3months} = 0.3$, $RE_{6months} = 0.2$) and the SF-MPQ-2 ($RE_{3months} = 0.9$, $RE_{6months} = 0.3$), the Short-WORC was less efficient (r < 1) at all time points, except at 3 months for the SF-MPQ-2.

Discussion

Overall, this study indicates that the prospective evaluation of the Short-WORC provides evidence of construct validity similar to the extracted WORC scores after an analysis at 6 months. While the Short-WORC was sensitive to change over 6 months, it was less responsive than the other comparator shoulder PROs, and performed similar to generic questionnaires. Given limitations in our sample for longer term follow-up these results may be unstable. Therefore, we have not established any measure to be superior, and further investigation is required before the Short-WORC can be qualified as being a measure that is valid and sensitive to change for evaluating QoL in rotator cuff pathology.

The interim findings in this study demonstrated that the Short-WORC can be used to discriminate between subgroups of sex and health status. As shown in our prior work, the extracted Short-WORC can be used to discriminate between the two groups of sex and worker's compensation status (with or without worker's compensation).[6] Those findings demonstrated that females and those who are not receiving worker's compensation have worse outcomes.[6] Our study prospectively demonstrates, the Short-WORC can also be used to discriminate between the

subgroups of sex and health status (changed vs unchanged) which is consistent with prior literature that indicates patients who score lower on the GRoC often report lower outcomes of recovery.[16]

The findings in this study further indicate that the Short-WORC demonstrated moderate correlations with other PROs when assessed over different time points. As shown, the EQ-5D-5L and the ASES had weak correlations with the Short-WORC, which can be expected for 2 reasons. Firstly, we expect measures of generic health - quality of life, such as the EQ-5D-5L and the shoulder-specific quality of life to be less related, than two shoulder specific measures due to the difference in scope. Secondly, given the Short-WORC does not have the emotional domain present in the full WORC, and only covers 2 of the 5 domains, we question if the Short-WORC is still even a QoL measure. Rather, we have stated that it is a functional outcome measure.[2,6] Therefore, the constructs being assessed are different and may account for the lower correlations.

Additionally, the ASES showed to be weakly correlated with the Short-WORC, which was not anticipated. It can be speculated that the differences in content and measurement metrics (0–10 versus 0–3) might have contributed to these results. As shown in a study by Vincent el al.,[25] the two elbow-specific measures of the Patient-Rated Elbow Evaluation form and the patient-reported form of the American Shoulder and Elbow Surgeons Elbow Questionnaire which should have correlated higher, resulted in a lower correlation due to the different scoring scales and items. Therefore, these differences could have influenced the correlations obtained between the Short-WORC and the ASES.

In regard to longitudinal validity (sensitivity to change), we see that the Short-WORC showed a change after 6 months, but did not demonstrate a large enough effect size like the condition – specific ASES. Furthermore, the Short-WORC had a similar effect size and SRM to the

SST, which is only a region- specific outcome measure. While our prior work reported the extracted Short-WORC scores to have an effect size of 1.05 and SRM of 0.89 after 6 months, the sample consisted of all surgical patients. In our current study, the sample population consisted of a mixed intervention group, where participants received an intervention of either surgical, physiotherapy or injections. By having a heterogeneous sample, we hypothesize that this could have lowered the reported SRM and effect sizes of the Short-WORC. Samples that contain all surgical patients would have experienced a higher effect after 6 months, explaining the higher SRM and ES values in our prior work. Furthermore, another group[25] recently used the WORC and ASES for an all surgical patient population, and had values that were similar to our prior published work on the Short-WORC, where the baseline measurements for the WORC had a mean value of 39.6 at baseline and 77.5 at 3 months. Therefore, future studies should further investigate administering the WORC in a population of patients undergoing a variety of treatments to compare the findings of this study. This will provide insight into whether the Short-WORC is better suited for detecting change in surgical patients only or can still be used for a heterogeneous intervention group.

Furthermore, analyzing longitudinal validity was calculated through effect size and SRM, which measure the magnitude of the changed score and do not necessarily indicate the importance of the observed change. While commonly used coefficients, they do not determine the levels of change in the measure.[11] Therefore, performing an analysis such as investigating the area under a receiver operating characteristic curve would be better at differentiating between clients who have improved and who have not improved by a significant amount. Additionally, as this study was an interim analysis, we did not assess responsiveness or an MCID value. While this study sample at 6 months was not large enough to conduct such an analysis, future studies should investigate defining an MCID with higher sample sizes at 6 months, and if it is similar to 11%-13% of the WORC.

Limitations

In this study, only certain psychometric properties of the Short-WORC were assessed. Therefore, future studies should evaluate the responsiveness through determining the minimal clinically important difference quantitatively and the comprehension and construct clarity of the Short-WORC qualitatively. Furthermore, all of the studies to date have been conducted at specialty shoulder surgery clinics, therefore, assessment in populations of different contexts or that include a broader spectrum of RCD would clarify whether these measurement properties exist in multiple contexts of the disorder.

In this study, we used a GRoC scale that consisted of a 3-point scale that ranged from better, the same, or worse. While this scale can detect the change of a patient's recovery, the fewer points do not always accurately discriminate the degree of patient improvement.[27] A study by Kamper et al. reviewed that a GRoC of 11-points or 15-points is stronger at discriminating between varying degrees of improvement or deterioration. [26] More points have been proven to offer a better compromise between patient preference, adequate discriminative ability, and test-retest reliability. Therefore, future studies should investigate using a GRoC that contains more points when trying to verify if the Short-WORC can discriminate for health status.

References

1. MacDermid JC, Ramos J, Drosdowech D, Faber K, Patterson S. (2004). The impact of rotator cuff pathology on isometric and isokinetic strength, function, and quality of life. *J Shoulder Elbow Surg.* 13(6):593-598.

2. Dewan, N., MacDermid, J. C., MacIntyre, N., & Grewal, R. (2016) Reproducibility: Reliability and agreement of short version of Western Ontario Rotator Cuff Index (Short-WORC) in patients with rotator cuff disorders. *Journal of Hand Therapy*, 29, PP. 281–291

3. Kirkley A, Alvarez C, Griffin S. (2003) The development and evaluation of a disease- specific quality-of-life questionnaire for disorders of the rotator cuff: The Western Ontario Rotator Cuff Index. *Clin J Sport Med*. 13(2):84-92

4. Holtby R, Razmjou H. (2005). Measurement properties of the Western Ontario rotator cuff outcome measure: a preliminary report. *J Shoulder Elbow Surg*. 14(5):506- 510.

5. Furtado, R., & MacDermid, J. C. (2018). Clinimetrics: Short Western Ontario Rotator Cuff Index. *Journal of physiotherapy*. 65(1), 56

6. Dewan, N., MacDermid, J., MacIntyre, N., & Grewal, R. (2015). Validity and responsiveness of Short Version of Western Ontario Rotator cuff Index (Short-WORC) in patients with rotator cuff pathology. 24,162-166.

7. Michener, L. A., McClure, P. W., & Sennett, B. J. (2002). American Shoulder and Elbow Surgeons Standardized Shoulder Assessment Form, patient self-report section: reliability, validity, and responsiveness. *Journal of shoulder and elbow surgery*, *11*(6), 587-594.

8. Richards, R. R., An, K. N., Bigliani, L. U., Friedman, R. J., Gartsman, G. M., Gristina, A. G., ... & Zuckerman, J. D. (1994). A standardized method for the assessment of shoulder function. *Journal of Shoulder and Elbow Surgery*, *3*(6), 347-352.

9. Dworkin, R. H., Turk, D. C., Revicki, D. A., Harding, G., Coyne, K. S., Peirce-Sandner, S., ... & Farrar, J. T. (2009). Development and initial validation of an expanded and revised version of the Short-form McGill Pain Questionnaire (SF-MPQ-2). *PAIN®*, 144(1-2), 35-42.

10. Godfrey, J., Hamman, R., Lowenstein, S., Briggs, K., & Kocher, M. (2007). Reliability, validity, and responsiveness of the simple shoulder test: psychometric properties by age and injury type. *Journal of Shoulder and Elbow Surgery*, *16*(3), 260-2

11. Lippitt, S. B., Harryman, D. T., & Matsen, F. A. (1993). The shoulder: a balance of mobility and stability. *Rosemont: American Academy of Orthopaedic Surgeons*, 545-559.

12. Rabin R, de Charro F (2001). EQ-5D: a measure of health status from the EuroQol Group. *Ann Med*. 33: 337–343.

13. Kamper, S. J., Maher, C. G., & Mackay, G. (2009). Global rating of change scales: a review of strengths and weaknesses and considerations for design. *Journal of Manual & Manipulative Therapy*, *17*(3), 163-170.

14. Furtado R, MacDermid JC, Bryant DM, Faber K, and Athwal G. (2019) Evaluating the reproducibility of the short version of the Western Ontario Rotator Cuff Index (Short-WORC) prospectively. Journal of Shoulder and Elbow Surgeons (submitted)

15. Kirshner B, Guyatt G. A methodological framework for assessing health indices. J Chronic Dis. 1985;38(1):27-36.

16. Razmjou H, Davis AM, Jaglal SB, Holtby R, Richards RR. (2011) Disability and satisfaction after rotator cuff decompression or repair: A sex and gender analysis. BMC *Musculoskelet Disord*.12:66

17. Streiner DL, Norman GR. Health measurement scales. A practical guide to their development and use. 4th ed. Oxford: Oxford University Press; 2008.

18. Finch E, Brooks D, Stratford PW, Mayo NE. (2002) Physical rehabilitation outcome measures: A guide to enhanced clinical decision-making. 2nd ed. ed. Hamilton, Ontario: BC Decker

19. Juniper EF, Guyatt GH, Feeny DH, Ferrie PJ, Griffith LE, Townsend M. Measuring quality of life in children with asthma. (1996). *Qual Life Res*.5(1):35-46.

20. Liang MH, Larson MG, Cullen KE, Schwartz JA. (1985) Comparative measurement efficiency and sensitivity of five health status instruments for arthritis research. *Arthritis Rheum*. 28(5):542-7

21. Fitzpatrick R, Ziebland S, Jenkinson C, Mowat A, Mowat A. (1993) A comparison of the sensitivity to change of several health status instruments in rheumatoid arthritis. *J Rheumatol.* 20(3):429-36.

22. Husted JA, Cook RJ, Farewell VT, Gladman DD. (2000) Methods for assessing responsiveness: A critical review and recommendations. *J Clin Epidemiol*. 53(5):459-68

23. Cohen J. (1988). Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale, New Jersey.

24. Wessel, R. N., Wolterbeek, N., Fermont, A. J., Lavrijsen, L., van Mameren, H., & de Bie, R. A. (2018). Responsiveness and disease specificity of the Western Ontario Rotator Cuff index. *Journal of orthopaedics*, *15*(2), 337-342.

25. MacDermid, J., Bryant, D., Holtby, R and Razmjou, H. (2019). Arthroscopic versus mini open rotator cuff repair: A randomized control trial and meta-analysis. Submitted to Journal of Bone and Joint Surgery.

26. Kamper, S., Maher, C., & Mackay, G. (2009) Global Rating of Change Scales: A Review of Strengths and- Weaknesses and Considerations for Design. *Journal of Manual & Manipulative Therapy*. 17:3,163-170.

Chapter 4: Interpretation and content validity of the items of the numeric rating version Short-WORC to evaluate outcomes in management of rotator cuff pathology: A cognitive interview approach.

Abstract

Background: The shortened version of the Western Ontario Rotator Cuff Index (Short-WORC) is a patient reported outcome measure that evaluates quality of life (QoL) of patients with rotator cuff pathology. However, formal content validation of the full or Short-WORC has not been reported. This study aims to understand how 1) people interpret and calibrate responses to items on the Short-WORC and 2) compensatory strategies that might enhance function and thereby affect responses.

Methods: This study used cognitive interviewing, a qualitative methodology that focuses on the interpretation of questionnaire items. Patients with rotator cuff disorders (n=10), clinicians (n=6) and measurement students (n=10) were interviewed using a talk aloud structured interview that evaluated each of the 7 items of the Short-WORC. All interviews were recorded and transcribed verbatim by one researcher (R.F). Analysis was done through an open coding scheme using a previously established framework which classified issues into 6 themes: Comprehension, Inadequate response definition, Reference Point, Relevance, Perspective Modifiers and Calibration Across Items

Results: Overall, the items on the Short-WORC were well received by participants, relevant and easily understood, with minor variations in interpretation. The items of working above the shoulder (90%), compensating with the unaffected arm (88%) and lifting heavy objects (92%) were the most relevant to participants. The items of sleeping and styling were coded by the theme of calibration across items (19%) frequently. Gender was a perspective modifiers for the items on styling your hair (30%) as it was more consistently relevant and more commonly required a multiple task component for women (19%). Compensatory strategies of using the other arm, altered positioning, help from others (task-re allocation) and using assistive devices/resources were frequently mentioned by participants.

Conclusions: Therefore, the findings demonstrate that the Short-WORC is not cognitively complex, but varies with patient perspectives. Overall, the Short-WORC provides evidence of strong content validity when used for rotator cuff disorder patients.

Keywords: patient reported outcomes, rotator cuff disorders, content validity, quality of life

A version of this work will be submitted for journal publication: Furtado R, MacDermid JC, Bryant DM, Faber K, and Athwal G. BioMed Central Health and Life Outcomes (2019)

Introduction

Rotator cuff disorders (RCDs) include a spectrum of pathologies that can lead to shoulder pain, impairment and activity limitation.[1] While the spectrum of disorders vary, rotator cuff tears are a common problem in the current population. Rotator cuff tears are commonly associated with exposure to repetitive movements or strain.[1,2] The prevalence of tears increases with age, affecting more than 60% of patients who are over the age of 60 [3] and results in a reduced quality of life (QoL).[1]

Since the goal of surgery and rehabilitation is to improve QoL in patients, understanding the construct of QoL is critical when defining optimal treatments.[3,4] The previous version of the Western Ontario Rotator Cuff Index (WORC) developed by Kirkley et al, is one of the most validated disease-specific questionnaire to measure QoL in RCDs.[5] The WORC focuses on 5 domains; 1) pain and physical symptoms, 2) sports and recreation, 3) work, 4) lifestyle, and 5) emotions.[5] While it has been translated and validated in a variety of different languages[1,2], the WORC created challenges of patient response burden (time spent to answer questionnaire) and complexity. [1,2]

A shortened version of the Western Ontario Rotator Cuff Index (Short-WORC) was created to address these concerns.[7] The Short-WORC by Razmjou et al. contains seven items from the domains of work and lifestyle, focusing on the activity limitations that arise from RCDs. [1,4,6,7] Participants completed a visual analogue scale (VAS) to score their response.[7] In the present study, we have modified the responsiveness scale to a 0-10-point numeric scale. In accordance with previously published work, a reduction in response burden and increase in patient satisfaction occurs with the use of a numeric scale.[8] As discussed in our prior work validating the Short-WORC, the concept of QoL may not be fully retained in this abbreviated questionnaire. Nevertheless, in our preliminary studies and those of others, the Short-WORC has demonstrated measurement properties that are similar to the original WORC.[1,6] While the Short-WORC demonstrates equally strong psychometric performance when compared to the full WORC, these previously published studies extracted the Short-WORC items from the fully completed original WORC.[1,6] Therefore, the Short-WORC has yet to be validated as an independent assessment.

A fundamental aspect of validation is understanding the content validity of a questionnaire.[9] Content validity refers to the extent to which a measure represents all facets of a given construct.[10] According to the Food and Drug Administration's (FDA) guidelines[11], content validity can be assessed through conducting interviews that seek to evaluate 1) the clarity of the instructions, 2) the content of each item and 3) that the intended meaning of each item is easily interpreted by participants. Understanding the recall and response scales in the PRO are also evaluated in the process.[9,10] Therefore, this study primarily aimed to evaluate content validity by exploring how people interpret and calibrate responses to items on the Short-WORC. A secondary aim of this study was to understand how compensatory strategies may influence the way participants interpret and determine responses to the items.

Methods

Study Design

This study used a descriptive qualitative approach based on the principles of cognitive interviewing to explore participants' interpretations of specific words, constructs (variables that cannot be measured directly but are informed through other variables that are measurable) and phrases in the Short-WORC. This enables an understanding of how participants calibrate options when responding to the measure.[12] Cognitive interviewing uses semi-structured interviews, a talk out loud approach, and probes to understand how patients interpret and respond to items on a self-report questionnaire.[12] This allows a combination of concurrent (while answering the question) or retrospective (immediately after answering the item) answering, which gathers optimal data quality.[12] Participants were provided with a version of the Short-WORC, that had a numerical scale from 0 -10.

Setting and sample

Interviews were conducted in a small private room at the Hand and Upper Limb Clinical Research Laboratory, London, Canada. Patient and healthcare provider participants were recruited from St. Joseph's Health Care London and researchers were recruited from Western University (London, Canada). Participants who met the inclusion criteria were invited to participate in the study; greater than 18 years of age, can speak and read English and did not have another mental or

physical aliment that could would confound the shoulder injury or not allow them to be able to participate in the interview process.

Through purposeful sampling, we aimed to include perspectives of healthcare providers and recipients.[13] Therefore, patients(n=10), healthcare providers (n=6), and measurement students (n=10) were recruited. Patients who had received some treatment for their shoulder (n=6) and patients who had yet to be exposed to evaluation for their shoulder (n=4) were recruited to achieve a diversity of participant experiences. Participation of both men and women of varying age groups allowed for a diversity of experiences. Recruitment for interviews stopped when saturation of the responses was achieved in each of the three participant groups.[12] The study protocol was reviewed and approved by Lawson Healthcare and Western Research Ethics Board (WREB).

Data collection

Participants provided written informed consent prior to the interview. Interviews were conducted in English by one researcher (RF) and lasted between 40 and 60 minutes. All interviews were recorded on an encrypted tape recorder, and then transcribed verbatim.

The interview structure was informed by previously published work[12,19] and multiple discussions with the research team. Interviews focused on participants' interpretation of each individual item on the Short-WORC. Through the think out loud approach, participants were encouraged to express all their thoughts when responding to each item. Probes such as, "Can you define this word?" or "Can you provide me with an example?", were asked to further explore the rationale of participants' specific responses to each item. Participants described how they determined (calibrate) their responses.

Analysis

Descriptive statistics (age, sex, occupational status and diagnosis) were collected and are presented in Table 1. The original audio recordings were analyzed by the research team. Analysis of the recordings were done through a descriptive thematic analysis, consisting of open coding.[12] This allowed the scripts to be characterized by fragments, in order for relevant themes to be extracted, categorized and classified. Themes were then identified from the responses to each item. Findings were summarized with quotes and percentages as appropriate. A previously established coding

system[1] was used to classify issues that affect interpretation. The categories include: Comprehension/clarity (C), Perspective modifiers (PM), Reference point (RP), Calibration across items(CAI), Inadequate response definition (IR) and Relevance (R).

Table 1: Demographic results of study population (n=26)

Groups		Patients who have received treatment (n=6)	Patients yet to be exposed to treatment (n=4)	Clinicians (n=6)	Measurement Students (n=10)
Age (years)		76	44	34	31
Female (%)		67	50	50	50
Male (%)		33	50	50	50
Professional status (%)					
	Student	0	25	33	100
	Employed	33	50	67	0
	Unemployed	0	0	0	0
	Sick-leave	0	0	0	0
	Retired	67	25	0	0
Affected shoulder (%)					
	Left	33	25	N/A	N/A
	Right	50	25	N/A	N/A
	Bilateral	17	50	N/A	N/A
Diagnosis (%)					
	Shoulder pain	67	75	N/A	N/A
	Rotator cuff tear	67	25	N/A	N/A
	TSA	67	0	N/A	N/A
	OA	67	0	N/A	N/A

Results

Data analysis resulted in the categorization of 6 themes that describe the issues that participants had when interpreting items. This is further described and illustrated by participants' quotes, as demonstrated in Table 2.

Overall findings demonstrate that individuals interpreted items based on their personal situation, personality traits, biology, work roles and/or environmental factors, which was represented through the theme of perspective modifiers. Certain items such as, styling your hair or dressing, were more influenced by gender. While the genders found this item relevant to QoL, males assigned less importance to this item when compared to items such as, working above shoulder level. Additionally, perspective modifiers also influenced the relevance of doing work outside the house for participants who specified living in an apartment or having designated help prior to the injury. Therefore, item relevancy to the construct is influenced by a person's biological, environmental or social context.

Instructions on Short- WORC

The Short-WORC focuses on two different domains which have unique sets of instructions. When participants (n=26) were asked to read the instructions out loud, some comprehension (C) issues arose. Specifically, there was misinterpretation around some of the important words in the instructions. For example, when asked to define the word "past week", some participants confused this with "average week".

"Okay so here I have to answer and think about my lifestyle and what I do on an average day in the week, and how much my shoulder has been kinda affecting or altering those activities." – Measurement Student #1, female

Some participants also struggled with the interpretation of the word "lifestyle". When asked to define, the majority of participants associated lifestyle with only activities of daily living.

"Lifestyle is my activities during the day. Would the term activities of daily living be better suited instead of lifestyle, is this what you are asking?" - Clinician #4, female

1. How much difficulty do you have sleeping because of your shoulder?

Overall, this item was well understood by most participants and did not demonstrate comprehension issues. When asked to define the term "sleeping", frequent words such as: "at rest, relaxed at night and lying down" arose. The majority of participants (73%) considered this item to be relevant to their lifestyle, however, some suggested that it was only relevant if they slept on the injured shoulder.

"I always sleep on my left side so my right shoulder (injured) is fine at night."- Patient #8, female

Additionally, some participants (19%) described compensator strategies that allowed them to sleep better at night. Participants discussed themes of intense shoulder pain, which translated to modifications of their sleeping position to comfortably rest.

"*Since the surgery was on my left shoulder, my left shoulder was out, and so I mostly slept on my back*." – patient #4, male.

2. How much difficulty have you experienced with styling your hair because of your shoulder?

Comprehension of this item was generally clear to all participants. When asked to define the term "styling your hair", phrases of "grooming, blow drying, combing, brushing and using styling products" arose. Findings demonstrate this item to be relevant to QoL with the majority of participants (79%), however, the theme of perspective modifiers heavily influenced the interpretation.

Relevance was a gendered issue with this item. Men more often found this item to be irrelevant to their QoL as they had short hair or were bald (30%).

"I'm bald, I don't need to style my hair!" – Patient #5, male.

Furthermore, some participants (19%) used the item of sleeping, as a reference point to calibrate their response to the item of styling your hair. Additionally, some participants identified that styling their hair was critical to their QoL, and so needed to compensate with the uninjured arm, seek assistance or allot more time in their day for styling.

"*If I were to injure my shoulder, I would still style my hair …I would get someone (roommate) to just help me out if I needed a specific style."- Measurement Student #1, female.*

3. How much difficulty do you have dressing or undressing?

In general, definitions of "dressing or undressing" were rephrased as "putting on clothes, removing clothes, and getting ready", indicating no comprehension issues. Findings further indicated that participants (70%) strongly endorsed the relevance of this item to their QoL. Additionally, some participants (30%) calibrated their response to this item, based on their scores for the items of sleeping and styling their hair.

"*I would say my answer would be the same as styling my hair…if I chose 5 or 6 in question 2, then I would choose the same answer for question 3." – Measurement Student #2, male*

Furthermore, participants identified the importance of completing this task and the need to compensate to complete it. Strategies for compensation included: requiring assistance from a device or family member, increasing the allotted time for changing of clothes, or changing the types of clothing worn in order to decrease shoulder movement.

"I can't reach my back to put on my bra…that's why my husband helps me out." – Patient #6, female.

4. How much difficulty do you experience in daily activities about the house or yard?

Findings indicated some comprehension issues with this item, as participants would interchange the words "lifestyle" and "daily activities" often. When asked to define "daily activities", terms such as: "chores, work, school, and living style" frequently arose.

"*Yeah my daily activities are defined by my work and hobbies. My life is my job, family and other activities I do." – Patient #5, female.*

In contrast, definitions of "about the house or the yard" resulted in phrases of: "chores, eating, cleaning, cooking, gardening and yard /outdoors work". Only one participant initially misinterpreted the meaning of about the house or yard, and defined it as occupational labour that involves working outdoors.

"This means work outside of the house like employment that you get paid for or yard work. This is both inside or outside the house and external jobs…that's what I think"- Clinician #6, Female

Overall, participants (80%) identified this item to be relevant to their overall quality of life, but some (12%) were concerned with the phrasing of "work in the yard". Due to participants' living conditions, some did not require the need to do yard work, i.e. living in an apartment or having designated help prior to the injury.

"No, I do not do any yard work, my husband always does that." – Patient #8, female.

Additionally, some compensatory strategies were mentioned such as: seeking assistance from someone else to do their daily activities or modifying the time period or frequency of activities they participated in.

5. How much difficulty do you experience with working above the shoulder?

Overall, this item received positive feedback from participants (90%), as many identified this item to be a critical component of recovery. Some participants (10%) identified that they did not need to do much overhead reaching and therefore, found this item less important to their quality of life.

"*I have an office job, I don't need to raise my arms much." – Patient #8, female.*

Definitions of "working above the shoulder" included phrases such as: "overhead reaching, lifting above my head and raising my arms", indicating comprehension was generally good for this item. Participants frequently mentioned compensating strategies in order to continue to work above shoulder level, such as: modifying the placement of items for easier access or seeking assistance when needing to reach above shoulder level.

"I try to use my left hand a lot more to help out and then I keep things within reach. The shelves are much lower in my house and if something is too high for me I use a step ladder." – Patient #10, female.

6. How much do you use your unaffected arm to compensate for the injured arm?

Definitions of "compensate" led to phrases such as: "using my not injured shoulder and using my healthy shoulder more", indicating no comprehension issues amongst participants. Furthermore, this item was identified as a critical component for QoL by participants (88%).

"I use my left hand a lot, which is much harder since I am very right-hand dominant." – patient #6, female.

In contrast, some participants (12%) indicated that compensating was less relevant, as their injury was on their non-dominant arm.

"I am right handed; my injury was on my left shoulder…do I compensate? Not frequently".

– Patient #3, female.

7. How much difficulty do you experience lifting heavy objects at or below shoulder level?

When asked to define "heavy objects", participants stated words such as: "weight, large and using force", indicating the item was well understood. Overall, this item resulted in a mix of responses depending on what stage of recovery the participant was in. Participants who were further along their recovery scored this item lower, while participants who were in the early stages of the injury scored it higher. Nevertheless, the majority of participants (92%) identified this item to be important to QoL. While evident that participants understood this item, some (12%) participate in a variety of tasks below shoulder level and therefore, were unsure which tasks to calibrate their score to.

"I do some yard work and cleaning that can be difficult to bend and pick up things from time to time...I think something in the middle?"- Patient, #3, female

Additionally, some participants discussed compensatory strategies such as re-allocating the task to someone else in order to feel less discomfort.

"Now with the snow coming, I will have to shovel myself since the weather is bad and I will have to find help."- Patient #7, female.

Discussion

Overall, the content validity of the Short-WORC was supported, as most respondents found the items on the Short-WORC to be clear and relevant to their functioning. However, the item of styling your hair was not relevant to a minority of the study sample and had a gender-bias being less relevant to men. Furthermore, it was evident that many patients had developed compensatory strategies as this was mediated in the difficulty reported. Overall, the items received positive feedback, there was no struggle with the recall period and most of the items were correctly interpreted.

The primary methodology in this study was cognitive interviewing. Cognitive interviewing in the health sciences was developed in the 1980s to improve questionnaires, as it evaluates the sources of response error.[18] This method draws on cognitive theory, which aids the understanding of how participants process and respond to questionnaire items, as there is a complexity of reasons behind each given response.[18] According to the Survey Interaction Model[18], cognitive processing alone does not account for all participant responses, as there is always underlying psychological processes such as emotions, personality, and biological characteristics.[18] For this reason, the application of cognitive interviewing to the initial pilot testing of PROs can help anticipate these factors during item selection. Furthermore, cognitive interviewing addresses the assumption that responses to items represent a common understanding of item content and intent across participants, allowing for the data to be used in further quantitative analyses in subsequent psychometric testing.[22]

According to the **CO**nsensus-based **S**tandards for the selection of health **M**easurement **IN**struments (COSMIN), content validity is one of the most important criterion for evaluating a PROs, and should be assessed by both patients and professionals. In 2018[15], COSMIN defined the standards for adequate content validity to be a measure that is comprehensive, comprehensible and relevant.[15,16] When evaluating a shortened PRO, it is advised that while some properties can be obtained from the original study, the property of comprehensiveness should be evaluated from a new study of the shortened PRO.[15] Therefore, it is important that researchers re-evaluate the content validity of a shortened PRO, to verify that it measures the intended construct of the original. This is

particularly important when the original development papers did not report detailed content validation, as is the case with the full WORC.

Comprehension of the items is also an important component of content validation according to the International Society for Pharmacoeconomics and Outcomes Research (ISPOR). They recommend that an item must use terminology that is consistent with its intended meaning and is relevant for respondents.[24] Findings demonstrate that only a few participants struggled with comprehension of the items, and the remaining were able to correctly define the items. Indicating that the comprehension level of the Short-WORC is appropriate for the intended population. Furthermore, aligning with the recommended guidelines by the ISPOR of using language that is understandable to all potential respondents, regardless of background.[13] Therefore, it is recommended to researchers to use appropriate language that does not diverge from the intended meaning. [13,19]

In addition, another facet of construct validation is item relevancy. Items must be relevant to the intended population and construct being assessed.[12] While results demonstrated a high percentage of relevancy to participants, none of the items were relevant to every individual participant. As anticipated, individuals are unique and have different opinions of what they calibrate as relevant to their QoL or recovery. While researchers try to anticipate this issue in the development of PROs, variables such as: gender, age, lifestyle, or social status will always hinder the relevancy of an item, as they are not generalizable.[13] In the original iteration of the WORC, Kirkley et al [5] used factor analysis and semi-structured interviews to rank the relevancy of items when measuring QoL. Therefore, researchers should use a variety of analytical methods during item selection to enhance the relevancy of items to the construct and the participants.

Selecting a responsive scale is another important criterion according to FDA, ISPOR and COSMIN guidelines.[15,16] Within this study, the Short-WORC was modified to a numeric scale from a visual analogue scale (VAS), as we found in prior pilot studies that patients were more comfortable responding on a 0-10 scale than a VAS. Participants responded favourably to the 0-10 scale, as it provided them an efficient way to accurately rate themselves. This modification reduces response burden, and clinician scoring burden, and is consistent with current studies that advocate the use of numeric scales. [17,18] Ultimately, psychometric data can be used to determine the optimal

response scale, as there is no universal standard.[22] We first reported the reliability and validity of a numeric Short-WORC extracted from a full WORC[1] and then reported on the reliability of it administered as a 7-item numeric scale.[27] This study establishes the content validity of the numeric version of the Short-WORC and specifically the ease of interpretation of the 0-10 scale.

The evaluation of participants' recall period is another component of cognitive interviewing. The recall period, assesses participants' responses based on the strategies they use when responding to an item.[12] The results confirm the lack of difficulty participants had in the recall phase when responding to the items, and there was no further indication of any unclear reference boundaries that could have impaired their responses. Furthermore, participants incorporated reference standards such as temporal anchors, to recall and calibrate their responses to. Additionally, some participants anchored their current response to an item based on the previous item response. This is evident due to the structure of the Short-WORC, such that the ordering of the items allows participants to easily recall their previous item's score as an anchor for the latter.[13] Therefore, confirming the structure of the Short-WORC to be clear and logical for participants; an important component of content validity.[12]

Based on the evidence presented in this study, we conclude that the Short-WORC achieved content validity. The findings demonstrated that all items were relevant to majority of participants when evaluating their QoL, and the issues were relatively minor. While a minority of participants found the item of *styling your hair* to be not relevant to their lifestyle, the majority of participants did. Changing existing measures is a major undertaking since it creates confusion and makes it less possible to compare data across time. Therefore, major issues should be present to warrant these changes. According to COSMIN guidelines[15], there is no reason to remove the item, but certain words could be replaced to improve clarity. Overall, it is evident that the comprehension levels of the Short-WORC were easy enough for all participants, and no major comprehension issues were identified that would result in the removal of items. Finally, it is evident that the recall period was accurately evaluated and participants found no difficulty in that process. Therefore, it is evident that the Short-WORC is a valid tool in regard to its content, and qualitatively validates our prior work.[1,6,7]

A secondary purpose of this study was to explore the compensatory strategies that influence the way participants calibrate their responses. Findings demonstrated strategies of modifying activity levels, altering positioning, lowering personal expectations or re-allocating tasks. Reallocation of tasks including timing (putting task off to later), using paid services (e.g. lawn maintenance or snow removal) and using members within their social support networks. The motivators for compensation included to reduce pain, to compensate for weakness or to not stress the shoulder given concerns about post-operative recovery. As shown in a study by Bialocerkowski[24] involving compensatory strategies with wrist problems, many people re-allocate tasks to family members to avoid additional stress to the affected limb, similar to the findings in our study. Understanding the spectrum of compensatory strategies for RCDs, provide further insight into why participants with similar impairments might report different levels of functional ability. From a treatment perspective sharing patient compensatory strategies may provide avenues to improve recovery or quality-of-life during recovery. Conversely, understanding compensatory mechanisms can provide insight into potential for other injuries.[24] For example, overcompensation with the uninjured arm, may increase the risk for an injury in the uninjured arm. Therefore, understanding the compensatory strategies is important to understand responses on PRO, functional recovery and future risks. Few tools consider compensation, and so the inclusion of this as an item on the WORC is an important consideration.[24]

Overall, limitations in this study included the use of a population that was predominately middle-class Caucasian. However, the demographic from this region where the study was conducted is predominately of Caucasian descent. [20,23] Therefore, future studies should gather information from other ethnic groups in order to compare and contrast the QoL. Furthermore, while the WORC was designed to measure quality of life, the Short-WORC focuses on activity limitation. Therefore, this study cannot be taken as supporting that the WORC and Short-WORC have concurrent content validity.

Conclusion

In conclusion, the evidence in this study demonstrates that there is no need for change to the items of the Short-WORC, as they are well-understood by patients with rotator cuff disorders. The Short-WORC accurately reflects the principles of comprehension, relevance and recall, therefore,

achieving content validity. Overall, the Short-WORC and items are an accurate measure of quality of life for rotator cuff pathology. Future studies should assess other psychometric properties such as reliability, validity and responsiveness prospectively.

Table 2: Common issues that arose with the Short-WORC. (n=26)

Item	Comprehension	Perspective Modifiers	Reference Point	Calibration Across Items	Inadequate Response definition	Relevance
1. Sleeping	Understood well					Overall, found to be relevant (73%) Few participants felt it was only relevant if the injured arm was slept on.
2. Styling your hair	Understood well	Gender issues arose (30%). E.x. men who were bald found this inapplicable		Some participants calibrated this answer with the item of sleeping (19%)		Overall, relevant to 79% of participants
3. Dressing or undressing	Understood well	Some gender issues arose, i.e. men who did not rate this item as important (19%)		Some participants calibrated this answer with the item of sleeping and styling hair		Overall, relevant to 70% of participants
4. Daily activities about the house or yard	Some comprehension issues with defining daily activities and yard work					Overall was relevant for 80% of participants 12% found work in the yard irrelevant to their Qol
5. Working above the shoulder	Understood well					Overall, 90% of participants found it was relevant to their QoL
6. Compensate with unaffected arm	Understood well					88% of participants found this relevant to their QoL. 12% found if the injury was on the non-dominant hand it was not relevant.

7.	**Lifting heavy objects at or below shoulder level**	Understood well	92% of participants found this to be relevant to QoL

References

1. Dewan N, MacDermid JC, MacIntyre N, Grewal R. Reproducibility: Reliability and agreement of short version of Western Ontario Rotator Cuff Index (Short-WORC) in patients with rotator cuff disorders. J Hand Ther. 2016 Jul-Sep;29(3):281-91.

2. Puga, Vanessa O. O., Lopes, Alexandre D., & Costa, Leonardo O. P. (2012). Assessment of cross-cultural adaptations and measurement properties of self-report outcome measures relevant to shoulder disability in Portuguese: a systematic review. *Brazilian Journal of Physical Therapy*, *16*(2), 85-93.

3. Johanson, M. E. (2016). Rehabilitation after surgical reconstruction to restore function to the upper limb in tetraplegia: a changing landscape. *Archives of physical medicine and rehabilitation*, *97*(6), S71-S74.

4. Michener LA, Synder, AR. Evaluation of health-related quality of life in patients with shoulder pain: are we doing the best we can? *Clin Sports Med*. 2008; Jul;27(3):491-505

5. Kirkley A, Griffin S. Development of disease-specific quality of life measurement tools. *Arthroscopy*. 2003;19(10):1121-8.

6. Dewan N, MacDermid JC, MacIntyre N. (2018).Validity and Responsiveness of the Short Version of the Western Ontario Rotator Cuff Index (Short-WORC) in Patients With Rotator Cuff Repair. *J Orthop Sports Phys Ther.* 48(5):409-418. doi:10.2519/jospt.2018.7928.

7. Holtby R, Razmjou H. (2015) Measurement properties of the Western Ontario rotator cuff outcome measure: a preliminary report. *J Shoulder Elbow Surg*. 14(5):506-510.

8. Lee, J. J., Lee, M. K., Kim, J. E., Kim, H. Z., Park, S. H., Tae, J. H., & Choi, S. S. (2015). Pain relief scale is more highly correlated with numerical rating scale than with visual analogue scale in chronic pain patients. *Pain physician*, *18*(2), 195-200.

9. Rentz AM, Skalicky AM, Burslem K, Becker K, Kaschinski D, Esser D, Revicki DA. (2017) The content validity of the PSS in patients with plaque psoriasis. J Patient Rep Outcomes.1(1):4. doi: 10.1186/s41687-017-0004-7.

10. Lawshe, C.H. (1975). A quantitative approach to content validity. *Personnel Psychology*, *28*, 563–575.

11. Food Drug Administration (FDA). (2009). Guidance for industry patient-reported outcome measures: use in medical product development to support labeling claims. Federal Register, 74(235), 65132–65133

12. Willis, G. B. (2004). *Cognitive interviewing: A tool for improving questionnaire design*. Sage Publications.

13. Damman, O. C., Hendriks, M., Rademakers, J., Delnoij, D. M., & Groenewegen, P. P. (2009). How do healthcare consumers process and evaluate comparative healthcare information? A qualitative study using cognitive interviews. *BMC Public Health*, *9*(1), 423.

14. Magasi, S., Ryan, G., Revicki, D., Lenderking, W., Hays, R., Brod, M., . . . Cella, D. (2012). Content validity of patient-reported outcome measures: Perspectives from a PROMIS meeting. *Quality of Life Research, 21*(5), 739-746.

15. Myrvik, M. P., Drendel, A. L., Brandow, A. M., Yan, K., Hoffmann, R. G., & Panepinto, J. A. (2015). A comparison of pain assessment measures in pediatric sickle cell disease: visual analog scale versus numeric rating scale. *Journal of pediatric hematology/oncology*, *37*(3), 190-194.

16. Murtagh, F. E., Addington-Hall, J. M., & Higginson, I. J. (2007). The value of cognitive interviewing techniques in palliative care research. *Palliative Medicine, 21*(2), 87-93.

17. MacDermid, JC. (2018) Cognitive interviewing (CI) - to identify sources of interpretation dissonance in in patient-reported outcome measures (PRO) [online]

18. Mercier, C., Peladeau, N., & Tempier, R. (1998). Age, gender and quality of life. *Community mental health journal*, *34*(5), 487-500.

19. Arumugam, V., MacDermid, J. C., & Grewal, R. (2013). Content analysis of work limitation, Stanford presenteeism, and work instability questionnaires using international classification of functioning, disability, and health and item perspective framework. *Rehabilitation research and practice.*

20. Statistics Canada (2016) Census Profile, 2006 Census. [website] Retrieved from: https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/details/page.cfm?Lang=E&Geo1=CMACA&Code1=555&Geo2=PR&Code2=35&Data=Count&SearchText=London&SearchType=Begins&SearchPR=01&B1=All&GeoLevel=PR&GeoCode=555&TABID=1

21. Thorne, S. (2016). *Interpretive description*. Routledge.

22. Wild, D., Grove, A., Martin, M., Eremenco, S., McElroy, S., Verjee-Lorenz, A., & Erikson, P. (2005). Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: report of the ISPOR Task Force for Translation and Cultural Adaptation. *Value in health*, *8*(2), 94-104.

23. Packham, T. L., & MacDermid, J. (2014). Content Validity of the Patient-Report Hamilton Inventory for Complex Regional Pain Syndrome. *Journal of Hand Therapy*, *27*(3), e3.

24. Bialocerkowski, A. E. (2002). Difficulties associated with wrist disorders–a qualitative study. *Clinical rehabilitation*, *16*(4), 429-440.

25. Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., ... & de Vet, H. C. (2010). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of clinical epidemiology*, *63*(7), 737-745.

26. Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., ... & De Vet, H. C. (2010). The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Quality of life research*, *19*(4), 539-549.

27. Furtado R, MacDermid JC, Bryant DM, Faber K, and Athwal G. (2019) Evaluating the reproducibility of the short version of the Western Ontario Rotator Cuff Index (Short-WORC) prospectively. *Journal of Shoulder and Elbow Surgeons* (submitted)

Chapter 5: General Discussion

Overall, this study provides evidence that the Short-WORC is a reliable, valid PRO which can detect sensitivity to change in patients with rotator cuff pathology. While prior work evaluated the Short-WORC scores when extracted from the full WORC, this work demonstrated that the Short-WORC was successful when assessed in a prospective RCD population. The results indicate that the Short-WORC is successful in retaining the same psychometric properties of the WORC. Despite some limitations with sample size and analysis, the Short-WORC still proves to be the best outcome measure for rotator cuff disorders like its parent, the WORC.

Strengthens and Limitations

While Kirkley et al.[1] used patient feedback to create the original items of the WORC, the results were not formally published. Therefore, the novel part of this study was using cognitive interviewing to validate the content of the Short-WORC. Cognitive interviewing in the process of PRO development allows for direct patient input. As demonstrated, the 6 themes used in this study can help researchers anticipate issues that may arise during outcome measurement development.[2,3]The most common themes that arose during the interview process were comprehension, perspective modifiers and relevance which seemed to heavily influence patient calibration of the Short-WORC items. Furthermore, cognitive interviewing is also useful during the translation of PROs, as shown in the systematic review of the cultural adaptions of the WORC. Cognitive interviewing helped translators successfully adapt each item to their appropriate culture, through the "expert committee and pre-testing" steps.[4] Through incorporating feedback from patients and clinicians, we enhance the likelihood of creating items that reflect our intended patient population; but also, limit issues that could hinder their understanding of an item.[2,3]

As its parent, the Short-WORC was intended to measure the construct of quality of life in patients with rotator cuff disorders. However, as Razmjou et al.[5] reduced the items through a factor analysis, only the domains of lifestyle and work remained for the new questionnaire. It can be argued that the Short-WORC measures the construct of activity limitation rather than QoL, as it does not include the emotional component of the original WORC. As shown, QoL incorporates both

the mental and physical components during recovery. Therefore, this study cannot be taken as supporting that the WORC and Short-WORC have concurrent content validity, and future studies should explore this limitation.

As shown in the literature, disease-specific measures gather explicit information related to a specific pathology; however, in cases where functional status of the shoulder is compromised by the existence of multiple pathologies, it can be challenging in determining the appropriate disease-specific measure.[6] Furthermore, the advantage of using rotator-cuff measures has not been established in the literature. While this work supports the Short-WORC as the most appropriate choice for RCDs, there is still further work that needs to be done in determining a gold standard for RCDs. Currently, in the literature there is no gold standard for shoulder outcome measures and so future research should investigate the criteria for determining and establishing a gold standard.

The patient response burden of the WORC suffered challenges that led to subsequent item reduction. Patient response burden includes complexity of a measure but also, administrative burden for the researcher or clinician. The goal of shortening the questionnaire is to reduce patient and/or administrative burden while retaining the conceptual linkage to the intended construct and sufficient psychometric equivalence.[5] Although we did not directly measure the time taken to complete the measure, we assumed that patients required less time to complete the 7-items of the Short-WORC compared to the 21-items of the original WORC. Furthermore, the questionnaires were administered on paper copies in clinic. While this assures a patient and clinician interaction, it can add to the administrative burden for the clinician. Many researchers have begun the shift towards electronic copies of questionnaires to limit missing data and administrative burden. However, before making the shift towards computerized PROs, it is important to evaluate if there is a difference between the scores obtained from paper compared to scores obtained electronically. A study by Godfrey et al., [7]compared the WORC scores when administered either electronically or with paper forms in a clinic waiting room. Results showed there was no difference in the WORC scores when administered in either way, however, found an increase in accessibility, ease and accuracy of recorded data when collected on a computer.[7] Therefore, future studies could investigate the advantages and disadvantages of using electronic copies of the Short-WORC, compared to the current method of administration.

Integrated Knowledge Translation and Clinical Implication

Overall, this thesis aimed to engage both researchers and knowledge users (orthopedic surgeons and physiotherapists) within each study of this thesis. These groups were critical in informing the development and execution of the research studies, and so will help to be engaged in sharing information about implementation of this PRO in the evaluation of RCDs in clinical practice and research.

Clinical implications would include the routine use of the Short-WORC in clinics that deal with RCD patients as this PRO has good clinical utility with no cost and can be administered with minimal training. Furthermore, the Short-WORC exhibited acceptable psychometric properties which would foster clinician confidence upon the results and interpretations obtained. Overall, the Short-WORC can aid clinicians in evaluating, discriminating and predicting activity limitation for rotator cuff pathology.

Future Directions

Based on our observations from the systematic review of the WORC, we recommend that studies should state a clear hypothesis when conducting psychometric studies, and make sure they follow recommended guidelines when translating items. We also recommend future studies include anchor based methods to calculate minimal clinically important difference and standard error of measurement, in order to clinically discriminate.

Conclusion

The Short-WORC is a reliable tool for patients with rotator cuff pathology.

References

1. Kirkley A, Alvarez C, Griffin S. The development and evaluation of a disease- specific quality-of-life questionnaire for disorders of the rotator cuff: the Western Ontario Rotator Cuff Index. Clin J Sport Med. 2003 Mar;13(2):84-92.

2. Willis, G. B. (2004). *Cognitive interviewing: A tool for improving questionnaire design*. Sage Publications.

3. MacDermid, JC. (2018) Cognitive interviewing (CI) - to identify sources of interpretation dissonance in in patient-reported outcome measures (PRO) [online]

4. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007;60 :34-42.

5. Holtby R, Razmjou H. Measurement properties of the Western Ontario rotator cuff outcome measure: a preliminary report. J Shoulder Elbow Surg. 2005 Sep- Oct;14(5):506- 510.

6. Dewan, N., MacDermid, J. C., MacIntyre, N., & Grewal, R. (2016). Reproducibility: reliability and agreement of short version of western Ontario rotator cuff index (short-WORC) in patients with rotator cuff disorders. Journal of Hand Therapy, 29(3), 281-291

7. Godfrey, Jenna M., MD, MSPH, Gines, J. L., PhD, & McCarty, E. C., MD. (2013). Comparison of computerized and paper versions of the western ontario rotator cuff (WORC) index.*Journal of Shoulder and Elbow Surgery*, *22*(4), 500-504.

Appendix 1. Short-WORC Questionnaire

SECTION A: Lifestyle

The following section concerns the amount that your shoulder problem has affected or changed your lifestyle. Please indicate the appropriate amount for the past week by circling a number.

1. How much difficulty do you have **sleeping** because of your shoulder?

0 1 2 3 4 5 6 7 8 9 10

2. How much difficulty have you experienced with **styling your hair** because of your shoulder?

0 1 2 3 4 5 6 7 8 9 10

3. How much difficulty do you have **dressing or undressing**?

0 1 2 3 4 5 6 7 8 9 10

SECTION B: Work

The following section concerns the amount that your shoulder problem has affected your work around or outside the house. Please indicate the appropriate amount for the past week by circling a number.

1. How much difficulty do you experience in **daily activities** about the house or yard?

0 1 2 3 4 5 6 7 8 9 10

2. How much difficulty do you experience **working above** the shoulder?

0 1 2 3 4 5 6 7 8 9 10

3. How much do you use your unaffected arm to **compensate** for the injured arm?

0 1 2 3 4 5 6 7 8 9 10

4. How much difficulty do you experience **lifting heavy objects at or below** shoulder level?

0 1 2 3 4 5 6 7 8 9 10

Appendix 2: Ethics

Western Research

Date: 15 June 2018

To: Dr. Joy MacDermid

Project ID: 111572

Study Title: The use of cognitive interviewing to evaluate the Short-WORC and the SANE in rotator cuff disorders

Application Type: HSREB Initial Application

Review Type: Delegated/Full Board

Full Board Reporting Date: July 3, 2018

Date Approval Issued: 15/Jun/2018

REB Approval Expiry Date: 15/Jun/2019

Dear Dr. Joy MacDermid

The Western University Health Science Research Ethics Board (HSREB) has reviewed and approved the above mentioned study as described in the WREM application form, as of the HSREB Initial Approval Date noted above. This research study is to be conducted by the investigator noted above. All other required institutional approvals must also be obtained prior to the conduct of the study.

Documents Approved:

Document Name	Document Type	Document Date
Cognitive Interview Questions (1)	Interview Guide	Received May 20, 2018
letter of information and consent V2	Written Consent/Assent	22/May/2018
SHORT – WORC	Paper Survey	
shoulder core HULC	Paper Survey	
The use of cognitive interviewing - Msc Thesis proposal	Protocol	Received May 22, 2018

No deviations from, or changes to, the protocol or WREM application should be initiated without prior written approval of an appropriate amendment from Western HSREB , except when necessary to eliminate immediate hazard(s) to study participants or when the change(s) involves only administrative or logistical aspects of the trial.

REB members involved in the research project do not participate in the review, discussion or decision.

The Western University HSREB operates in compliance with, and is constituted in accordance with, the requirements of the TriCouncil Policy Statement: Ethical Conduct for Research Involving Humans (TCPS 2); the International Conference on Harmonisation Good Clinical Practice Consolidated Guideline (ICH GCP); Part C, Division 5 of the Food and Drug Regulations; Part 4 of the Natural Health Products Regulations; Part 3 of the Medical Devices Regulations and the provisions of the Ontario Personal Health Information Protection Act (PHIPA 2004) and its applicable regulations. The HSREB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 00000940.

Please do not hesitate to contact us if you have any questions.

Sincerely,

Karen Gopaul, Ethics Officer on behalf of Dr. Joseph Gilbert, HSREB Chair

Note: This correspondence includes an electronic signature (validation and approval via an online system that is compliant with all regulations).

Appendix 3. Curriculum Vitae

Rochelle Furtado

Education

Western University – Masters of Science in Health and Rehabilitation with a specialization in Physical Therapy M.Sc.

London, Ontario 2017- present

- Research in the field of Orthopaedics, Measurement outcomes and Upper Limb extremities. Supervisor: Dr. Joy MacDermid
- Expected Graduation – August 2019

University of Ottawa – Honours Bachelor of Science with a specialization in Molecular and Cellular Biology B.Sc.

Ottawa, Ontario 2017

- Honour's Project **-** Quantitative analysis of muscle atrophy in the rabbit supraspinatus after complete tendon tear: effect of re-attachment. Supervisor: Dr. Odette Laneuville

Funding and Awards

- 2019 – Student Presenter Grant - Student Faculty Showcase - $200.00
- 2018 - Bone and Joint Trainee Award - $ 500.00 annually
- 2018 - Graduate Student Conference Travel Award - $125.00
- 2017, 2018- Western Research Scholarship - $ 3000.00 annually
- 2013, 2014 & 2015 – Student Bursary Award D&H LTD. - $300.00
- 2013 - University of Ottawa Admission Scholarship - $1000.00

Research Experience

Summer Research Student

*Hand and Upper Limb Clinical Research Lab***,** *Lawson Research Institute*

2017-present

- Conducted quantitative research projects with emphasis on patient recruitment, testing various measurements outcomes on participants and analyzing data via SPSS software.
- Conducted qualitative research projects with emphasis on interviewing participants and analyzing data with proper theoretical underpinnings.

- Completed ethics applications (initial and recommendations) for on-going HULC research projects.

Summer Research Student

Bone and Joint Lab, *University of Ottawa Faculty of Medicine* 2016

- Honour's research project entitled: Quantitative analysis of muscle atrophy in the rabbit supraspinatus after complete tendon tear: effect of re-attachment.
- Created a protocol that can quantify muscle atrophy, by exploring the change in skeletal myocyte cross sectional area in different experimental conditions post tendon tear.
- Collected data in a histology lab with emphasis on microscopy, analysis with ImageJ software and statistics with SPSS software.

Publications and Papers in Progress

Published/In press/Accepted

1. **Furtado, R.,** & MacDermid, J. C. (2019). Clinimetrics: Short Western Ontario Rotator Cuff Index. *Journal of physiotherapy*. 65(1) pg 56
2. **Furtado, R**., & MacDermid, J. C. (2019). Clinimetrics: Single Assessment Numeric Evaluation. *Journal of physiotherapy*. (In press)
3. Bobos, P., Nazari, G., **Furtado, R.** and MacDermid, J. (2019). A Systematic Review of the Minimally Clinically Important Differences in Grip Strength in Patients with Musculoskeletal and Neurological Conditions (Accepted- *Physiotherapy Canada*)
4. **Furtado, R.,** MacDermid, J., Nazari, G., Bryant, D., Faber, K. and Athwal, G. (2019) Cross-cultural adaptions and measurement properties of the WORC (Western Ontario Rotator Cuff Index): a systematic review (Accepted – *BMC Health and Quality of Life Outcomes*)

Under Review

1. Nazari, G, MacDermid, J., Bobos, P. and **Furtado, R.** (2019). Psychometric Properties of the Single Assessment Numeric Evaluation (SANE) in Patients with Shoulder Conditions. A systematic review. (under review -Orthopaedics & Traumatology: Surgery & Research)
2. **Furtado, R.,** MacDermid, J., Bryant, D., Faber, K. and Athwal, G. (2019). Evaluating the reproducibility of the short version of the Western Ontario Rotator Cuff Index (Short-WORC) prospectively (Under review- Journal of Shoulder and Elbow surgery)
3. **Furtado, R.,** MacDermid, J., Nazari, G., Bryant, D., Faber, K. and Athwal, G. (2019) Cross-cultural adaptions and measurement properties of the Shoulder Pain and Disability Index (SPADI): a systematic review (under review– *Journal of Hand Therapy*)

Conferences and Presentations

Poster Presentations

1. **Furtado, R.,** MacDermid, J., Bryant, D., Faber, K. and Athwal, G. Evaluating the content validity of the short version of the Western Ontario Rotator Cuff Index (Short-WORC) prospectively. London Health Sciences day – London, 2019
2. **Furtado, R.,** MacDermid, J., Bryant, D., Faber, K. and Athwal, G. Evaluating the content validity of the short version of the Western Ontario Rotator Cuff Index (Short-WORC) prospectively. Health and Rehabilitation Sciences Graduate Symposium – London, 2019
3. Farzad,M., MacDermid, J. and **Furtado, R.** Abilities and Barriers to Life Engagements (ABLE). MacHAND Day' 18 – Hamilton, 2018
4. **Furtado, R.,** MacDermid, J., Bryant, D., Faber, K. and Athwal, G. The use of cognitive interviewing to evaluate the Short-WORC and the SANE in rotator cuff pathology- Canadian Bone and Joint Conference '18 - London, 2018
5. **Furtado, R.,** Lanueville, O., Trudel, G. and Uhthoff H. Quantification of muscle atrophy in rabbit supraspinatus muscle. University of Ottawa Undergraduate Biology/Biomedical Science day- Ottawa, 2017

Reviewer

- **Abstract Reviewer.** Health and Rehabilitation Sciences Graduate Symposium – London, 2019
- **Abstract Reviewer.** Congress Montreal18 – Montreal, 2018
- **Abstract Reviewer/Volunteer.** Health and Rehabilitation Sciences Graduate Symposium – London, 2018

Moderator

- **Moderator.** *Oral Sessions A-C*. Health and Rehabilitation Sciences Graduate Symposium – London, 2018
- **Moderator.** *Session 3 Mental Health with Dr. Bhatla*. University of Ottawa Health Care Symposium – Ottawa, 2017
- **Moderator.** *Session A*. University of Ottawa's Regenerating New Ideas: A focus on the Future of Regenerative Medicine - Ottawa, 2015

Volunteer

- **Volunteer.** Health and Rehabilitation Sciences Graduate Symposium – London, 2019
    - Member on the marketing committee and designed all conference programs and media advertising.
- **Volunteer.** Health and Rehabilitation Sciences Graduate Symposium – London, 2018

    - Signed in delegates at the registration desk and helped for set-up of event.
- **Volunteer.** University of Ottawa Health Care Symposium – Ottawa, 2017
    - Assisted in promotion on social media and ticket sales for the event, and helped in set-up of the event.
- **Delegate/Volunteer**. An evening with the Canadian Association of Regenerative Medicine – Ottawa, 2015
    - Signed in delegates at the registration desk and helped for set-up of event.

Invited Presentations

- **Ethics in practise.** OT 9541, Western University, December 2018

Editorial Experience

- **Journal Reviewer.** Journal of Hand Therapy (n=4)
- **Journal Reviewer.** Health Science Inquiry (n=3)

Internal Extracurricular Services

- **HRS Representative.** Bone and Joint Training and Leadership Committee 2019- present
    - Participated in outreach events to promote BJI to the community
    - Contributed to ideas in improving trainee learning at BJI
- **HRS Representative.** Communications Committee, Bone and Joint Training and Leadership Committee 2019 – present
    - Assisted in running social media accounts for promotion of BJI
- **Vice President of Student Development**. Graduate HRS Student Council 2018-2019
    - Held monthly events to promote student development for graduate students, i.e., presentation workshop, thesis writing workshop etc.
- **General Member**. University of Ottawa Health Care Council 2017
    - Participated in events and helped in social media promotion for events

External Extracurricular Services

- **President**. Initiative Club, Our Lady of Mount Carmel Secondary
  2012-2013
    - Lead group meetings and hosted events to promote social justice issues at large.

- **Vice-President**. Initiative Club, Our Lady of Mount Carmel Secondary
  2011-2012
    - Lead group meetings and hosted events to promote social justice issues at large.

- **Co-President.** Student Council St. Richard School
  2008-2009
    - Initiated group meetings and hosted social events

Certifications

- Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans Course on Research Ethics (TCPS 2: CORE) 2018
- (AODA) Excelling at Accessible Customer Service 2018
- Standard Operating Procedures– Lawson Research Institute 2018
- Workplace Hazardous Materials Information System (WHIMS) 2018
- First AID/CPR 2015
- Completed Grade 8 Piano under the Royal Conservatory of Music 2013

Work Experience

Graduate Teaching Assistant September 201 –Present

- Graded essays, projects and held office hours for courses OT 9541 and HS2000
- Mentored students in course content for HS 2000 and OT 9541
- Proctored midterm and final examinations for various undergraduate courses in the Department of Biology and Psychology.

Disability Tax Credit Consultant May – June 2017

- Met weekly sale targets of bringing in new clients
- Assisted eligible candidates in the process of applying for the disability tax credit
- Communicated with various clients via phone and email to help resolve issues and concerns

Pedalheads Bike Camp Counsellor May 2015- August 2015

- Coached younger children in outdoor biking activities and road safety lessons.

- Planned weekly themed activities to promote wellness and physical activity.
- Communicated students' success with parents though report cards and daily discussions at the end of camp.

Academic Tutor - Our Lady of Mount Carmel Highschool 2011-2013

- Assisted students in Grade 9- 10 for Math and Science; afterschool for 2 hours a week per subject.
- Developed lesson plans with weekly quiz questions to test students' progress.
- Guided students with efficient strategies on how to better prepare for tests and complete assignments.
- Communicated with teachers on different methods and lesson plans to engage students

Volunteer Experience

St. Joseph's Healthcare London – Hand and Upper Limb Physiotherapy Department 2018- present

- Observed physiotherapists administering different treatments for patients.
- Assisted in floor tasks such as changing beds, tidying rooms and gym area for patients.
- Participated in 1 on 1 discussions with physiotherapist on patient cases.

St. Joseph's Healthcare London – Chronic Pain Physiotherapy Department 2018- present

- Patient interaction by providing hot packs and resistance bands for their exercises.
- Administrated work such as faxing and filing patient charts and documents.
- Assisted in floor tasks such as changing beds, tidying rooms and gym area for patients.

The Ottawa Hospital Volunteer 2015 - 2017

- Assembles the patient charts with discretion and distributes health surveys for hospital research study
- Conducts phone calls to patients in regard to upcoming clinics.
- Administrated work such as faxing and filing patient charts and documents.
- Catalogued Patient Progress Reports, and then mailed them to the respective family doctor

Canadian Association for Stem Cell Research Member 2014-2015

- Assisted at symposiums promoting the awareness of stem cell research in the hospital to the students and faculty of the University of Ottawa.
- Assisted in ticket sales and general venue set up for the symposium
- Attended general meetings once a month to collaborate new ideas with other members for future events.

Camp counsellor for the iCanBike program July 2014

- Coached children with musculoskeletal disabilities how to ride two-wheeler bikes by themselves or with training wheels.
- Motivated children by promoting physical outdoor activities tailored to their strengthens
- Assembled lesson plans for parents to accommodate their children at home

Skills and Abilities Field Related

Laboratory/ Technical skills

- Experienced in both qualitative and quantitative research methods – HS 9601, 9602
- Certified in TCPS and SOP through Lawson Health Institute
- Demonstrated microscopy skills (i.e. animal model dissections - Animal Form and Function Bio 2135, acquisition of images and utilizing software such as ImageJ and Infinity Capture.
- Familiarized in molecular biology lab techniques such as Polyacrylamide gel electrophoresis and double digestions with restriction enzymes in E. coli - Molecular Biology Bio3170.
- Familiarized with cell biology techniques such as immunohistochemistry with polycolonal and monoclonal antibodies in order to recognize epitopes and fluorescence microscopy techniques. i.e. DAPI
- Computed statistical analysis using software such as R program and SPSS.

Communication Skills

- Recruited patients for a hospital performance research study
- Corresponded with clinicians and researchers at the Bone and Joint Research Lab and at the Ottawa General Hospital

- Demonstrated strong presentation skills acquired through weekly laboratory meetings and Fourth Year Biology seminar course - BIO 4920,4921
- Presenter of guest speakers at the University of Ottawa Health Care Symposium 2017
- Attended the Ottawa General Health and Safety in Patient Care - Corporate Training. 2015

Teaching Skills

- GTA position for the faculty of science and health science at Western University
- Accompanied Grade 5 students as a student volunteer for a class field trip to the Ontario Science center - 2015
- Guided first year university students to the faculty of science in campus tours and through social events to help connect students during 101 Week 2016.
- Mentored incoming Grade 9 students through the LINKCrew program that allows a successful transition into highschool through monthly discussions, social activities and retreats.
- Volunteered at Community Life service club at the University of Ottawa for the promotion and teaching of French culture to other students