ARTICLES

A Shortened Version of the Western Ontario Rotator Cuff Disability Index: Development and Measurement Properties

Helen Razmjou, MSc, PT, PhD;[*†] *Paul Stratford, MSc, PT;*[‡] *Richard Holtby, MD, FRCSC*[*§]

ABSTRACT

Purpose: The objective of this study was to develop and examine the measurement properties of a shortened version of the Western Ontario Rotator Cuff Index (WORC), the SHORTWORC, in individuals with rotator-cuff pathology. ***Methods:*** The study occurred in two stages, both using secondary analysis of existing data sets. The first stage used cross-sectional data from candidates for rotator-cuff surgery to develop the SHORTWORC. The second stage examined various measurement properties of the SHORTWORC by analyzing scores from the WORC, the American Shoulder and Elbow Surgeons questionnaire, and the Constant–Murley score obtained from patients before and after rotator-cuff surgery. Approaches to validating the SHORTWORC included calculating the standard error of measurement (SEM) at an instant in time, performing a confirmatory factor analysis, correlating findings among the questionnaires, and examining differences between men and women. Sensitivity to change was investigated using standardized response mean and relative efficiency. ***Results:*** Data for 712 patients were used to develop the SHORTWORC, the final version of which consisted of 7 questions. Data for 166 patients (86 men, 80 women; mean age 57 $\pm$ 11 years) were used for validation. The SEM based on internal consistency (SEM_{IC}) was calculated as 7.43 SHORTWORC points. The SHORTWORC had similar convergent validity ($r = 0.72 - 0.82$) and sensitivity to change (SRM = 1.20 vs. 1.25, $p > 0.05$) to the longer version. The relative efficiency of the SHORTWORC was 3.19 times that of the WORC (95% CI, 1.50–71.51) in discriminating men's from women's level of disability. ***Conclusions:*** The SHORTWORC has indicators of validity, relative efficiency, and sensitivity to change comparable to those of the original version but has a smaller response burden.

Key Words: outcome measures; reliability of results; shortened version; validity of results; WORC.

RÉSUMÉ

Objectif : L'objectif de cette étude était de créer et d'analyser les propriétés de mesure d'une version abrégée du *Western Ontario Rotator Cuff Index* (WORC), le SHORTWORC, chez les personnes aux prises avec une pathologie de la coiffe des rotateurs. ***Méthode :*** Ce projet s'est déroulé en deux étapes. Une analyse secondaire des ensembles de données existants a été réalisée à chacune d'elles. La première a fait appel à des données transversales de candidats à une intervention chirurgicale pour la coiffe des rotateurs pour l'élaboration de l'indice abrégé SHORTWORC. La deuxième étape s'est penchée sur les diverses propriétés de mesure de l'indice SHORTWORC en analysant les pointages obtenus dans le cadre du WORC, les résultats au questionnaire de l'*American Shoulder and Elbow Surgeons* et les indices de l'échelle Constant-Murley; tous ont été obtenus auprès des patients avant intervention chirurgicale pour la coiffe des rotateurs. Les outils de validation du SHORTWORC comprenaient notamment l'erreur type de mesure (ETM) à un point précis dans le temps, l'analyse du facteur de confirmation, la corrélation des conclusions des questionnaires et l'examen des différences entre les hommes et les femmes. La sensibilité au changement a également été analysée à l'aide de la moyenne des réponses pondérées et de l'efficacité relative. ***Résultats :*** Des données pour 712 patients ont été utilisées pour l'élaboration de SHORTWORC, dont la version finale comportait 7 questions. Des données relatives à 166 patients (86 hommes et 80 femmes, âge moyen de 57 $\pm$ 11 ans) ont été utilisées aux fins de validation. L'erreur type de mesure basée sur la constance interne SEM_{IC} a été calculée de façon à représenter 7,43 points SHORTWORC. Le SHORTWORC avait une validité convergente similaire ($r = 0,72 - 0,82$) et une sensibilité au changement (ETM = 1,20 comparativement à 1,25, $p > 0,05$) avec la version plus longue de l'indice. L'efficacité relative du SHORTWORC était 3,19 fois plus importante que le WORC (95% IC: 1,50–71,51) au moment de distinguer le degré d'incapacité des homme de celui des femmes. ***Conclusions :*** Le SHORTWORC possède des indicateurs de validité, d'efficacité relative et de sensibilité au changement comparables à la version originale de l'indice, et constitue un moins lourd fardeau au chapitre des réponses à fournir.

From the: *Holland Orthopaedic & Arthritic Centre, Sunnybrook Health Sciences Centre; †Department of Physical Therapy; §Department of Surgery, University of Toronto, Toronto; ‡School of Rehabilitation Science, Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ont.

Correspondence to: Helen Razmjou, Holland Orthopaedic and Arthritic Centre, 43 Wellesley St. E., Toronto, ON M1Y 1H1; helen.razmjou@sunnybrook.ca.

Contributors: All authors designed the study, collected the data, and analyzed and interpreted the data; drafted or critically revised the article; and approved the final draft.

Competing Interests: None declared.

Acknowledgements: Dr. Alexandra Kirkley was a visionary orthopaedic surgeon who developed several disease-specific shoulder measures in the late 1990s, when certain methodological concepts were still in their infancy. Dr. Kirkley passed away prematurely, but her legacy lives on through her tremendous contribution to promoting high standards in the research and clinical practice of orthopaedic surgery in Canada.

Physiotherapy Canada **2012; 64(2);135–144; doi:10.3138/ptc.2010-51**

"Rotator cuff" is a collective term for four tendons: supraspinatus, infraspinatus, teres minor, and subscapularis. Failure of the rotator-cuff tendons due to wear and tear is the most common clinical problem of the shoulder,[1] accounting for more than 4.5 million physician visits per year in the United States[2] and substantial work loss.[3–5]

Several measurement tools exist for patients with shoulder problems. Upper-extremity measures are designed to be sensitive to disorders affecting any part of the upper limb;[6–9] shoulder-specific outcome measures have been developed to examine the impact of a condition on a specific joint;[10,11] and disease-specific shoulder measures focus on aspects of health that tend to be affected by a specific disease or condition.[12–17]

The Western Ontario Rotator Cuff Index (WORC), a disease-specific outcome measure, was developed by researchers at the University of Western Ontario in London, Canada, approximately 10 years ago.[16] The developers of the WORC[16] criticized previous shoulder measures for poorly defining techniques for physical examination, weighting items arbitrarily, failing to consult with patients, and using double-barrelled questions, but they were not specific about which measures had which shortcomings. Their rationale for developing a disease-specific shoulder measure was a need for measures of health-related quality of life (QOL) to assess the benefits of orthopaedic interventions.[14–16] Thus, the developers of the WORC consider this questionnaire a disease-specific QOL measurement tool for rotator-cuff pathology.[16] It should be noted that the appropriateness of labelling rotator-cuff dysfunction as a "disease" is debatable.

The first step in developing the original WORC involved a review of the literature and of existing measurement tools. Based on the World Health Organization's definition of health as "a state of complete physical, mental and social well-being,"[18(p.1)] the developers felt that five domains would be appropriate: physical symptoms, sports and recreation, work, lifestyle, and emotions. The second step was to accumulate items from discussions with health care professionals with an interest in shoulder pathology. The third step involved interviewing 30 patients (23 male, ages 30–76 y). Item reduction was achieved by eliminating duplicated, incomprehensible, or ambiguous items. The reduced list of 76 items was then administered to a randomly selected sample of 100 patients, after which the investigators eliminated questions that were highly correlated. The resulting 21 individual items, which used a visual analogue scale (VAS), were then given to 10 patients for further evaluation of wording.

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, as in instability and rotator-cuff pathology, it is difficult to determine which disease-specific instrument is the most appropriate. In addition, the advantage of rotator-cuff measures has not been proven to date in the shoulder-related literature.[19–21] Although the main value of the multiple-domain measures is in their ability to document the impact of disease on each QOL domain, we do not have supporting evidence for the specificity of information from each domain or sub-scale.[22,23]

Several other shortcomings specific to the WORC can be noted. The major concerns relate to the lack of support for factorial validity of the five domains and the illogic of a single summative score that reflects these domains. Since a total score should represent a common theme, rather than a compilation of different concepts, a multidimensional tool should not produce a single summative score. Moreover, when one purpose of a measure is to assess change over time, the items contributing to the total or sub-scale score should display similar change trajectories; this is not the case when assessing pain versus function, as the two recover differently. The WORC's response burden (i.e., the effort required to respond) adds to the challenge of collecting information in busy clinical settings. At present, the superiority of this lengthy multiple-domain disease-specific measure over shorter versions and joint- or limb-specific measures remains the subject of controversy. Therefore, developing a shorter version of the WORC (SHORTWORC) that minimizes the response burden of the original but preserves confidence in the scores is desirable.

The objectives of this study were (1) to develop a shortened version of the WORC based on the literature and clinical judgment and (2) to investigate the reliability, validity, and sensitivity to change of this tool. We hypothesized that the identified items of the SHORTWORC would form a single construct. We also hypothesized that the shorter version would have a moderate to high ($r = 0.50–0.80$) cross-sectional convergent validity with other competing shoulder questionnaires; that the SHORTWORC would be able to differentiate between men's and women's levels of disability with a similar relative efficacy; and, finally, that it would demonstrate similar sensitivity to change as the original version.

METHODS

Our study used two data sets from patients operated on over a 10-year period by one surgeon specializing in shoulder reconstruction in an academic centre. The cross-sectional data were used for item reduction; the cohort data were used to evaluate measurement properties of the shorter version. Approval for use of the existing databases was obtained from the Research Ethics Board of the Sunnybrook Health Sciences Centre.

The inclusion criterion for the initial stage of the study (cross-sectional data) was the presence of rotator-cuff pathology requiring surgical treatment (acromioplasty for impingement syndrome, tendinitis, or partial-thickness tears, or rotator-cuff repair for full-thickness tears). Patients with associated pathologies of the biceps requiring debridement, tenodesis, or tenotomy were included. Ex-

clusion criteria were previous surgery and associated additional pathologies that required other major surgeries, such as stabilization or superior labral repair. Patients with minor pathology in the rotator-cuff tendons underwent arthroscopic decompression with or without lateral resection of the clavicle. Patients with full-thickness tears of the rotator cuff underwent arthroscopic or open repair of the tendon(s). Some procedures overlapped (e.g., some patients underwent both repair and acromioplasty). All patients whose data were included in analysis completed the WORC questionnaire 2 to 3 weeks before surgery.

Stage 1: Development of the SHORTWORC

On traditional scales of highly correlated items, items are usually manifestations of an underlying hypothetical construct (*effect indicators*). QOL or disability indexes include a less homogenous set of items, such as symptoms or side effects of an illness (*causal indicators*). Questionnaires composed of effect indicators, such as depression or anxiety inventories, can be evaluated through traditional psychometric techniques, which are based on the assumption that all items of the scale reflect the latent construct that the scale is designed to measure. On questionnaires composed of symptoms and functional difficulties, however, the items do not necessarily reflect a common latent factor and can act independently. Most commonly, symptoms are strongly related to other QOL factors such as emotional or social functioning.

Streiner[24] and Fayers and Hand[25,26] have provided comprehensive reviews of how statistical analyses of these two types of questionnaires differ. Streiner has noted[24] that correlations among effect indicators are due solely to the items' relationships with the underlying construct. For causal indicators, however, covariances may or may not exist among items, irrespective of their relationship with the construct.[24–26] The specific items may not be related to one another, and some may in fact be mutually exclusive (e.g., clicking in the shoulder and difficulty with overhead lifting). The magnitude of the correlations may therefore change radically from one population to another. In an example relevant to our study, Streiner explained that rotator-cuff pathology may cause difficulty in reaching behind to do up a bra yet not interfere with the ability to put on stockings or a skirt.[24]

Fayers and Hand have noted that "the role of Exploratory Factor Analysis (EFA) should be restricted to situations in which there are unlikely to be symptoms or other causal items which may affect QOL,"[26(p.149)] highlighting that (1) "EFA cannot model indicators which have a causal effect upon QOL";[26(p.146)] (2) "EFA cannot assign weights to causal indicators";[26(p.147)] (3) "EFA may frequently extract strange combinations of causal indicators as factors";[26(p.147)] and (4) "EFA will yield different factors in different patient subgroups, according to the treatment and disease under investigation."[26(p.147)] They advised that "confirmatory factor analysis (CFA), which tests the goodness of fit of a pre-specified factor model, is considered ... to be a far more appropriate method for construct validation, on theoretical grounds."[26(p.139)] A recent article on methodological approaches to assessing the quality of outcome studies further highlighted the differences between reflective models (composed of effect indicators) and formative models (composed of causal indicators).[27]

Because statistics such as coefficient alpha, mean interitem correlation, and EFA are based on the assumption of item homogeneity and may not be fully appropriate for questionnaires composed of causal indicators, development of the SHORTWORC was guided by theoretical and clinical principles.

The WORC index places significant emphasis on symptoms and activity limitations (e.g., difficulty lifting, dressing and undressing, styling hair) that are not necessarily representative of QOL. Although the relevance of these questions to different domains of QOL is questionable, individual items reflect common attributes of patients with different levels of rotator-cuff pathology and cover activity limitations commonly seen in this population. Prior to designing the shortened measure, we decided that it should assess a single concept consistent with the International Classification of Functioning, Disability, and Health (ICF) framework that considers impairment, activity, and participation.[28] Previous literature on factor loading of the original WORC[23] and higher discriminant validity of the work and lifestyle domains of the original version[29] suggested a common theme focused around the importance of difficulty in performing activities.

Nine of the original WORC items include the word "difficulty," and one other item considers "ability" (i.e., ability to throw). Complementing these items is an item that appears to comment indirectly on functional status ("How much use of uninvolved arm?"). Our second step was a detailed qualitative examination of each of these items. We eliminated items that do not represent regular activities (e.g., push-ups, throwing, contact with shoulder); we also removed the item containing "roughhousing," because many patients whose data were used in our study did not feel that this item was applicable to them. This left seven items, including all items from the work and lifestyle domains except the "roughhousing" question. The SHORTWORC thus evaluates activity limitations rather than health-related QOL. We then used a CFA of the original cross-sectional data to specify a one-factor measurement model with uncorrelated error terms. The final step was to examine the extent to which the factorial structure and internal consistency of these seven items were supported in the cross-sectional data.

Stage 2: Cross-validation of the SHORTWORC

Our goal in the second stage was to examine measurement properties of the SHORTWORC in a different sample. Cross-validation analysis used prospectively collected data for patients with similar pathologies who

were involved in another formal study. This data set did not include data that could be used to assess test–retest reliability, minimal detectable change, or minimal clinically important change. All patients had been administered the WORC, along with with two shoulder-specific outcomes, the Constant–Murley score (CMS) and the American Shoulder and Elbow Surgeons questionnaire (ASES), 2 to 3 weeks before surgery and at 6 months after surgery. A standardized rehabilitation programme appropriate to the type of surgery was given to all patients, to be performed under physiotherapist supervision.

Outcome measures

Western Ontario Rotator Cuff Index (WORC)[16]

As noted, the WORC has 21 questions divided into 5 domains: (1) physical symptoms (6 questions); (2) sports and recreation (4 questions); (3) work (4 questions); (4) lifestyle (4 questions); and (5) emotions (3 questions). The total raw score, determined by measuring 21 lines (0–10 mm) that document pain or difficulty on a VAS, may range from 0 to 2100; lower scores are associated with fewer symptoms. The final score is reported as a percentage derived by subtracting the raw total from 2100, dividing by 2100, and multiplying by 100; a higher final percentage is associated with fewer symptoms. The updated (1998) version of the WORC, which was used with the patients in our database and which differs slightly from an earlier version that was published in 2003, was obtained from the original authors in 2000. The differences between the versions have been highlighted by other investigators.[23] The WORC has been reported to be reliable[16] and valid[19,20,29,30] in patients with rotator-cuff disease.

Short Western Ontario Rotator Cuff Index (SHORTWORC)

The SHORTWORC has seven items, including all items from the WORC work and lifestyle domains except the one relating to roughhousing. The highest raw (most symptomatic) score on the SHORTWORC is 700; the best (asymptomatic) score is 0. The SHORTWORC score, reported as a percentage derived by subtracting the total from 700, dividing by 700, and multiplying by 100, varies between zero (most symptomatic) and 100 (asymptomatic).

As a rule of thumb, if answers to 10% of questions are missing for an index, the index is considered to be missing completely. Therefore, the SHORTWORC, with only seven items, has no allowance for missing data.

American Shoulder and Elbow Surgeons questionnaire (ASES)[11]

The ASES is a 100-point scale (0 = most symptomatic, 100 = no disability); 50 points are derived from patient self-report of pain on a VAS, and the other 50 are computed from a formula using the cumulative score of self-report of difficulty related to 10 activities of daily living, derived using a 4-point ordinal scale. The ASES has been reported to be reliable[31] and valid[20,21,31,32] in patients with upper-extremity, shoulder, or rotator-cuff pathology.

Constant–Murley Score (CMS)[10]

The CMS combines a subjective component (35% of the total score) with the objective clinical assessment of range of motion (ROM) and strength (65% of the total score). A unique feature of the CMS is its ability to convert the absolute score to the relative score by adjusting for age and sex, which increases its usefulness with older populations. The CMS is more costly than subjective measures, as it requires trained clinicians to perform the objective measures of ROM and strength. The CMS is reported to be reliable and valid in patients with shoulder- or rotator-cuff-related pathologies.[21,32,33]

Statistical analyses

Univariate analyses, including the mean, median, standard deviation, 25th and 75th percentiles (first and third quartiles), and full range of scores, were used to describe the distribution of scores for the long and short versions of the WORC. The initial step of the cross-validation process was to perform a CFA using AMOS software (SmallWaters Corp., Chicago IL). The Cronbach's alpha coefficient was used to test whether the SHORTWORC items represented a unidimensional concept, keeping in mind that factor loadings and error variances may vary in different samples, as is commonly seen in formative models. In addition, coefficient alpha was used to calculate the standard error of measurement based on internal consistency (SEM_{IC}) at an instant in time.[34] The SEM_{IC} is useful in interpreting the confidence in the total score; it is calculated as $SEM = SD_{pooled}\sqrt{1-\alpha}$, where SD_{pooled} is the pooled pre- and postoperative standard deviations. Cross-sectional *convergent validity* of the WORC and SHORTWORC was examined by correlating their scores with those of the ASES and relative CMS, which is adjusted for sex and age differences. Pearson's correlation coefficient was applied to examine convergent validity. To further examine construct validity, we investigated the SHORTWORC's ability to discriminate between men's and women's disability levels. We calculated the *efficiency* of the SHORTWORC relative to the WORC ($F_{SHORTWORC}/F_{WORC}$; a value of 1 indicates that the efficiency of the measures is identical). To examine *known-groups validity*, a form of construct validation determined by the degree to which an instrument can demonstrate different scores for groups known to vary on the variables being measured, we examined the ability of the WORC and SHORTWORC to differentiate between full-thickness tear and impingement syndrome / partial-thickness tear. A one-way analysis of variance (ANOVA) was conducted to answer these questions.

Sensitivity to change, a component of a measure's longitudinal validity, relates to the capacity of a measure to detect change over time.[35,36] We applied the standardized response mean (SRM)—calculated as the mean change score divided by the standard deviation of the change scores—to assess the relative abilities of the WORC and SHORTWORC to detect change. A larger

Table 1 Description of the Development and Validation Stages of the SHORTWORC

Stage 1 (development)	
Sample characteristics	
n	712
Type of data	Prospective historical cohort, cross-sectional analysis
Age, mean (SD) y	56(13); range: 20–90
Sex, M:F	445 (62.5%) : 267 (37.5%)
Type of surgery,* no. (%)	366 (51): Rotator-cuff repair 491 (69): Arthroscopic acromioplasty 262 (37): Resection of lateral end of clavicle
Methods	
Literature review	Review of the items and domains of relevant studies
Item reduction	Qualitative step: • Selecting items that focused on functional status • Eliminating items that did not represent regular activities Quantitative step: • CFA to confirm unidimensionality
Stage 2 (cross-validation)	
Sample characteristics	
n	166
Type of data	Prospective longitudinal data, secondary analysis
Age, mean (SD) y	57(11); range: 32–80
Sex, M:F	86 (52%) : 80 (48%)
Type of surgery,* no. (%)	110 (66): Rotator-cuff repair 141 (85): Arthroscopic acromioplasty 95 (57): Resection of lateral end of clavicle
Methods	
Hypothesis testing	CFA of the validation sample Internal consistency Convergent validity Known-groups validity Relative efficiency Sensitivity to change

*Some surgeries overlapped.
SHORTWORC = Shortened version of the Western Ontario Rotator Cuff Index; M = male; F = female; CFA = confirmatory factor analysis.

SRM represents a greater ability to detect change. Because confidence intervals (CIs) for the SRM cannot be calculated directly, we applied a bootstrap procedure: 1,000 paired bootstrap samples with replacement, each with a sample size of 166, were obtained for the WORC and SHORTWORC; the boundaries of the 95% CIs were identified as the 25th and 975th values. The difference in SRMs between WORC and SHORTWORC was obtained by subtracting the SHORTWORC value from the WORC value for each of the 1,000 paired SRMs. The 95% CI boundaries for the difference were identified as the 25th and 975th difference values. To provide context to assist in interpreting the magnitudes of the SRMs for the SHORTWORC and WORC, we also calculated SRMs for the ASES and CMS measures.

RESULTS

Cross-sectional data for 712 consecutive surgical candidates who provided complete responses to all questions and who had different levels of rotator-cuff pathology were used in developing the SHORTWORC. A cohort of 166 patients with similar pathologies who had complete scores for all outcome measures (WORC, ASES, and CMS) was used for cross-validation. Table 1 provides the sample characteristics and methods used for each stage.

Distribution

Mean and median preoperative scores for the WORC and SHORTWORC were similar in magnitude in the cross-sectional sample: mean (SD) 39 (18) and 39 (20); median 39 and 38; min/max 5/89 and 2/85; Q1 = 26 and 24; Q3 = 51 and 54. Postoperative data showed a similar distribution for WORC and SHORTWORC respectively: mean (SD) 70 (24) and 70 (25); median 77 and 78; min/max 10/99 and 0/100; Q1 = 52 and 51; Q3 = 90 and 91. The Kolmogorov–Smirnov test of preoperative data indicated similar properties for long and short versions (0.058, $p > 0.15$, and 0.06, $p > 0.15$, respectively). Correla-

Table 2 Demographics of the Cross-Validation Sample

Variables	No. (%) of patients $n = 166$
Sex	
Male	86 (51.8)
Female	80 (48.2)
Dominant side	
Left	12 (7.2)
Right	153 (92.2)
Ambidextrous	1 (0.6)
Affected side	
Left	48 (28.9)
Right	118 (71.1)
Symptoms	
Pain on movement	143 (86.1)
Night pain	121 (72.9)
Stiffness	69 (41.6)
Weakness	117 (70.5)
Clicking/catching	79 (47.6)
Mechanism of injury	
Insidious	62 (37.3)
Fall	25 (15.1)
Direct blow	5 (3.0)
Traction	4 (2.4)
Repetitive activities	32 (19.3)
Type of surgery*	
Repair	110 (66.3)
Acromioplasty	141 (84.9)
Resection of lateral clavicle	95 (57.2)
Size of full-thickness tear†	
Small	3 (2.7)
Moderate	83 (75.5)
Large	18 (16.4)
Massive	6 (5.5)
Biceps pathology	
Partial tear	15 (9.0)
Complete rupture	32 (19.3)
Subluxed/dislocated	3 (1.8)

*Some surgeries overlapped.
†Size of tear reported only in patients with full-thickness tears ($n = 110$).

tions between the pre- and postoperative SHORTWORC and WORC were high (Pearson product–moment correlation, $r = 0.94$ for preoperative and $r = 0.97$ for postoperative data), indicating comparability of scores between the full WORC and the short version. Further details of demographics and descriptive statistics of items of the sample used for cross-validation are shown in Tables 2 and 3. All outcome measures showed a statistically significant improvement in disability over a period of 6 months (see Table 3).

Coefficient alpha and standard error of measurement (SEM) at an instant in time

A coefficient alpha value of 0.89 indicates a high correlation among the items representing difficulty with activities in the sample studied. The SEM_{IC} was calculated as 7.43 SHORTWORC points.

Confirmatory factor analysis

The initial CFA of the original sample (712) specified a one-factor measurement model with uncorrelated error terms. Model fit was assessed by applying the Comparative Fit Index (CFI), Tucker–Lewis Index (TLI), and root mean square error of approximation (RMSEA). No singularly agreed-upon standard exists to define acceptable fit; however, it is generally accepted that CFI and TFI values >0.95 indicate a good fit; RMSEA values of 0.08–0.05 indicate reasonable fit, and RMSEA values <0.05 represent good fit.[37]

Table 4 displays the results of the CFA for the development and cross-validation samples. For the development sample, the initial analysis resulted in fit coefficients >0.97; however, the RMSEA was 0.117, larger than the recommended standard of 0.08. Based on the reported modification indexes, we specified a second model that allowed a correlation between the error terms for "Difficulty working above shoulder" and "Use of uninvolved arm" and between the error terms for "Difficulty styling hair" and "Difficulty dressing/undressing." The fit coefficients for the second model exceeded 0.99, and the RMSEA was 0.065. Similar findings were obtained for the cross-validation sample, with fit coefficients >0.96 and an RMSEA $= 0.083$ for the correlated error term model.

Construct validity

The SHORTWORC detected a difference in functional status between men and women ($F_{1,164} = 10.05$, $p = 0.002$), but the WORC ($F_{1,164} = 3.15$, $p = 0.07$) did not.

The SHORTWORC was 3.19 times as efficient as the WORC (95% CI, 1.50–71.51).

Cross-sectional convergent validity

The preoperative and 6 months postoperative cross-sectional convergent validity coefficients are presented in Table 5. Preoperative coefficients are displayed below the main diagonal; 6 months postoperative coefficients are shown above the main diagonal. The magnitudes of pre- and postoperative WORC and SHORTWORC correlation coefficients did not differ significantly when correlated with the ASES ($r_{difference} = 0.02$, $Z = 0.96$, $p = 0.33$) and CMS ($r_{difference} = 0.01$, $Z = 0.91$, $p = 0.36$).

Known-groups validity

Neither measure was able to detect a statistically significant difference between levels of pathology (SHORTWORC: $F_{1,164} = 0.11$, $p = 0.738$; WORC: $F_{1,164} = 0.01$, $p = 0.937$).

Sensitivity to change

The SRMs and differences in SRMs are presented in Table 6. The main diagonal reports the SRMs for each

Table 3 Descriptive Statistics for SHORTWORC Items in the Cross-Validation Sample ($n = 166$)

	Mean (SD) score; range	
Variable	Preoperative period	Postoperative period
Daily activity	59.89 (23.85);0.00–100.00	28.01 (27.08); 0.00–100.00
Working above shoulder	79.22 (18.99); 5.00–100.00	41.47 (33.20); 0.00–100.00
Compensate	68.03 (25.59); 5.00–100.00	37.78 (34.18); 0.00–100.00
Lifting heavy objects	52.51 (29.97); 0.00–100.00	30.92 (28.46); 0.00 –100.00
Sleeping	64.77 (27.10); 0.00–100.00	28.93 (29.83); 0.00–100.00
Styling hair	51.93 ((32.09); 0.00–100.00	24.08 (28.98); 0.00–100.00
Dressing and undressing	49.67 (25.25); 0.00–100.00	19.43 (23.20); 0.00–100.00
SHORTWORC*	39.13 (20.18); 2.14–85.00	69.91 (25.16); 0.00–100.00
WORC†	39.31 (17.61); 5.00–89.00	69.51 (23.74); 10.00–99.00
RCMS‡	47.40 (21.01); 2.17–98.55	79.97 (24.17); 8.21–123.18
ASES§	48.14 (19.93); 5.00–90.00	75.93 (20.16); 16.66–100.00

* $t = 15.42$; $p < 0.0001$.
† $t = 16.10$; $p < 0.0001$.
‡ $t = 17.10$; $p < 0.0001$.
§ $t = 14.74$; $p < 0.0001$.
SHORTWORC = shortened version of the WORC; WORC = Western Ontario Rotator Cuff Index; RCMS = Relative Constant–Murley score; ASES = American Shoulder and Elbow Surgeons questionnaire.

Table 4 SHORTWORC Items and Confirmatory Factor Analysis Coefficients

	Standardized coefficients			
	Development sample ($n = 712$)		Cross-validation sample ($n = 166$)	
Items	Uncorrelated error terms	Correlated error terms	Uncorrelated error terms	Correlated error terms
Work 11: How much difficulty do you experience in **daily activities** about the house or yard?	0.87	0.89	0.84	0.87
Work 12: How much difficulty do you experience **working above your shoulder**?	0.75	0.74	0.67	0.66
Work 13: How much do you use your uninvolved arm to **compensate** for your injured one?	0.69	0.67	0.64	0.63
Work 14: How much difficulty do you experience **lifting heavy objects** at or below shoulder level?	0.72	0.72	0.64	0.65
Lifestyle 15: How much difficulty do you have **sleeping** because of your shoulder?	0.66	0.66	0.71	0.71
Lifestyle 16: How much difficulty have you experienced with **styling your hair** because of your shoulder?	0.75	0.71	0.79	0.74
Lifestyle 18: How much difficulty do you have **dressing or undressing**?	0.77	0.74	0.81	0.77
Fit coefficients (CFI, TLI)	>0.975	≥0.991	≥0.93	≥0.96
RMSEA	0.117	0.065	0.107	0.083

SHORTWORC = Shortened version of the Western Ontario Rotator Cuff Index; CFI = Comparative Fit Index; TLI = Tucker–Lewis Index; RMSEA = root mean square error of approximation.

measure; the off diagonals present the difference in SRMs between the intersecting measures. The mean (SD) change scores for the WORC and SHORTWORC were 30.20 (24.17) and 30.8 (25.72) respectively. These values yielded SRMs of 1.25 (95% CI, 1.04–1.51) for the WORC and 1.20 (95% CI, 1.00–1.44) for the SHORTWORC. The difference in SRMs between measures was 0.05 (95% CI, −0.01–0.12) in favour of the WORC; however, the 95% CI included zero. For purposes of comparison, the SRM for the ASES was 1.14 (vs. 1.33 for the CMS). The difference in SRMs between the CMS and ASES was significant—the 95% CI of the difference did not include zero—

Table 5 Cross-Sectional Convergent Construct Validity Coefficients (95% CI)

	CMS	WORC	SHORTWORC	ASES
RCMS	—	0.82* (0.76–0.86)	0.81* (0.75–0.86)	0.82* (0.76–0.86)
WORC	0.66† (0.56–0.74)	—	0.97* (0.96–0.98)	0.84* (0.79–0.88)
SHORTWORC	0.64† (0.54–0.72)	0.93† (0.91–0.95)	—	0.83* (0.78–0.87)
ASES	0.67† (0.57–0.75)	0.75† (0.68–0.81)	0.77† (0.70–0.83)	—

*Post-operative coefficients.
†Preoperative coefficients.
RCMS = Relative Constant–Murley score; WORC = Western Ontario Rotator Cuff Index; SHORTWORC = Shortened version of the WORC; ASES = American Shoulder and Elbow Surgeons questionnaire.

Table 6 Standardized Response Means (95% CI) and Differences in Standardized Response Means

	RCMS	WORC	SHORTWORC	ASES
RCMS	1.33* (1.15–1.55)	—	—	—
WORC	0.08† (−0.15–0.28)	1.25* (1.04–1.51)	—	—
SHORTWORC	0.13† (−0.08–0.33)	0.05† (−0.01–0.12)	1.20* (1.00–1.44)	—
ASES	0.19† (0.01–0.35)	0.11† (−0.08–0.30)	0.06† (−0.13–0.24)	1.14* (0.96–1.38)

Note: The only statistically significant difference is between the CMS and ASES (0.19; 95% CI, 0.01–0.35).
*Difference in SRM.
†SRM.
RCMS = Relative Constant–Murley score; WORC = Western Ontario Rotator Cuff Index; SHORTWORC = Shortened version of the WORC; ASES = American Shoulder and Elbow Surgeons questionnaire; SRM = standardized response mean.

but the SRM of the CMS did not differ significantly from those of the WORC and SHORTWORC.

DISCUSSION

The results of our study indicate that the 7-item SHORTWORC produces scores with similar measurement characteristics to the 21-item original WORC. As noted earlier, certain statistics may not be applicable to questionnaires composed of items addressing symptoms and functional difficulties, such that items are not expected to be fully correlated at all times or in all patients. Internal consistency has no interpretable meaning in tools composed mostly of causal indicators. Streiner[38] has noted that, except for extremely narrowly defined traits, $\alpha > 0.90$ is likely to indicate unnecessary redundancy rather than a desirable level of internal consistency. He highlighted the fact that the blind use of coefficient alpha and other indexes of internal consistency, with no consideration of their appropriateness for the measure in question, can lead to a scale's being wrongly dismissed for lack of reliability or unfairly criticized for not yielding useful results. In our case, the coefficient alpha of 0.89 supports the premise that the items are accessing the same concept without redundancy among items. It is likely that other samples of patients with rotator-cuff pathology will present with a slightly higher or lower alpha, depending on the extent of their symptoms and disability. The SEM based on coefficient alpha is an error estimate specific to a single application or an instant in time. For any reported SHORTWORC score, there is a 68% chance (± 1 SEM) that the true score lies within ± 7.43 points.

In terms of content validity, we consider the SHORTWORC and WORC to be comparable with respect to functional difficulty, since all WORC items that relate to activity limitation are included in the SHORTWORC with the exception of one question about roughhousing (an activity to which most individuals with shoulder problems do not relate).

The high correlation between WORC and SHORTWORC suggests highly comparable scores. The SHORTWORC performed similarly to other shoulder questionnaires and was better able than the WORC to differentiate between men's and women's levels of disability. Sensitivity to change was also comparable between the two versions. Our findings also suggest that although the ASES and CMS used a less rigorous approach to development (based on clinician censuses vs. patient interviews), they are not necessarily less valid or reliable than a disease-specific measure, as the developers of the WORC suggested.[16] It has been suggested that upper-extremity measures have comparable measurement properties in patients with rotator-cuff pathology.[6–9] Further research and head-to-head comparison of the SHORTWORC with upper-extremity measures

of physical impairment are required to determine optimal methods for assessing outcomes following interventions to treat rotator-cuff disorders.

Long instruments have a significant time and cost burden, and the higher variability that results from the greater number of questions may make them less useful in research. Higher variability affects the standard deviation and error of measurement and thus influences the sample-size calculation for formal studies, particularly randomized controlled studies.

Relative to the full WORC, the SHORTWORC provides similar information on physical limitations secondary to rotator-cuff problems, at a lower cost to clinicians and researchers (number of pages to copy, time spent scoring the questionnaire) and a smaller response burden for patients (time spent answering the questionnaire). Future research should focus on developing consensus-based standards from an international group of experts to select the most valid, reliable, and efficient outcome measures for the shoulder joint.

Several limitations of our study should be noted. First, the study used data from two samples of patients operated on by one surgeon specializing in shoulder reconstruction in an academic centre. The applicability of our results may be limited to surgical candidates with chronic symptoms who choose to undergo surgery after conservative treatment has failed; non-surgical candidates undergoing conservative treatment may have a different response pattern. Second, because the research questions were tested on previously collected data, it was not possible to assess the SHORTWORC's test–retest reliability, minimal detectable change, minimal clinically important difference, or time to complete and score. Finally, the replacement of the VAS with the adjectival scale warrants further investigation. The main advantages of the VAS are its continuous format and avoidance of the imprecise descriptive terms used in some rating scales; however, VAS scales are time-consuming to score by clinicians, lose accuracy when photocopied, and require a higher level of concentration and coordination, which adds to complexity of their use. Our future work will investigate these issues.

CONCLUSION

The SHORTWORC is comparable to the WORC for descriptive and analytical purposes related to functional difficulty secondary to rotator-cuff pathology. Additional research is required to further investigate the lack of superiority of long, multi-domain, condition-specific questionnaires in assessing patient outcomes following shoulder surgeries.

KEY MESSAGES

What is already known on this topic

The literature has identified several problems with existing measures for rotator-cuff disease, including using a total score that represents a compilation of different concepts that do not display similar change trajectories; lack of specificity of information provided by each domain; difficulty in using a specific measure in the presence of multiple pathologies (e.g., rotator-cuff pathology and osteoarthritis) commonly seen in patients with shoulder problems; and the response burden of lengthy measures.

What this study adds

A shorter version of a disease-specific measure (the WORC) was developed and examined to demonstrate why certain statistics may not be fully applicable to questionnaires composed of symptoms and functional difficulties (referred to as *causal indicators*). As part of this work, we were able to show that the SHORTWORC provides similar information on physical limitations secondary to rotator-cuff problems, with a potentially lower cost for clinicians and researchers and lower burden for patients.

REFERENCES

1. McFarland EG, Kim TK, Park HB, et al., editors. Examination of the shoulder: the complete guide. New York: Thieme Medical Publishers; 2006.
2. Oh LS, Wolf BR, Hall MP, et al. Indications for rotator cuff repair: a systematic review. Clin Orthop Relat Res. 2007;455:52–63. doi:10.1097/BLO.0b013e31802fc175. Medline:17179786
3. Herberts P, Kadefors R, Högfors C, et al. Shoulder pain and heavy manual labor. Clin Orthop Relat Res. 1984;191(191):166–78. Medline:6499308
4. Zakaria D, Robertson J, MacDermid JC, et al. Estimating the population at risk for Ontario Workplace Safety and Insurance Board-covered injuries or diseases. Chronic Dis Can. 2002;23(1):17–21. Medline:11876832
5. Zakaria D. Rates of carpal tunnel syndrome, epicondylitis, and rotator cuff claims in Ontario workers during 1997. Chronic Dis Can. 2004;25(2):32–9. Medline:15554610
6. Beaton DE, Katz JN, Fossel AH, et al. Measuring the whole 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. J Hand Ther. 2001;14(2):128–46. Medline:11382253
7. Beaton DE, Wright JG, Katz JN, and the Upper Extremity Collaborative Group. Development of the QuickDASH: comparison of three item-reduction approaches. J Bone Joint Surg Am. 2005;87(5):1038–46. doi:10.2106/JBJS.D.02060. Medline:15866967
8. Davis AM, Beaton DE, Hudak P, et al. Measuring disability of the upper extremity: a rationale supporting the use of a regional outcome measure. J Hand Ther. 1999;12(4):269–74. Medline:10622192
9. Stratford PW, Binkley JM, Stratford DM. Development and initial validation of the upper extremity functional index. Physiother Can. 2001;53:259–67.
10. Constant CR. An evaluation of the Constant-Murley shoulder assessment. J Bone Joint Surg Br. 1997;79(4):695–6. Medline:9250770
11. Richards RR, An K, Bigliani LU, et al. A standardized method for the assessment of shoulder function. J Shoulder Elbow Surg. 1994;3(6):347–52. doi:10.1016/S1058-2746(09)80019-0.
12. Gill TM, Feinstein AR. A critical appraisal of the quality of quality-of-life measurements. JAMA. 1994;272(8):619–26. doi:10.1001/jama.272.8.619. Medline:7726894
13. Kirkley A, Griffin S, Dainty K. Scoring systems for the functional assessment of the shoulder. Arthroscopy. 2003;19(10):1109–20. doi:10.1016/j.arthro.2003.10.030. Medline:14673454

14. Kirkley A, Griffin S, McLintock H, et al. The development and evaluation of a disease-specific quality of life measurement tool for shoulder instability. The Western Ontario Shoulder Instability Index (WOSI). Am J Sports Med. 1998;26(6):764–72. Medline:9850776
15. Lo IK, Griffin S, Kirkley A. The development of a disease-specific quality of life measurement tool for osteoarthritis of the shoulder: The Western Ontario Osteoarthritis of the Shoulder (WOOS) index. Osteoarthritis Cartilage. 2001;9(8):771–8. doi:10.1053/joca.2001.0474. Medline:11795997
16. 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;13(2):84–92. doi:10.1097/00042752-200303000-00004. Medline:12629425
17. Hollinshead RM, Mohtadi NG, Vande Guchte RA, et al. Two 6-year follow-up studies of large and massive rotator cuff tears: comparison of outcome measures. J Shoulder Elbow Surg. 2000;9(5):373–81. doi:10.1067/mse.2000.108389. Medline:11075319
18. World Health Organization. Constitution of the World Health Organization. Suppl. to: Basic documents. 45th ed. Geneva, Switzerland: 2006 [cited 2011 Nov 16]. Available from: http://www.who.int/governance/eb/who_constitution_en.pdf.
19. MacDermid JC, Drosdowech D, Faber K. Responsiveness of self-report scales in patients recovering from rotator cuff surgery. J Shoulder Elbow Surg. 2006;15(4):407–14. doi:10.1016/j.jse.2005.09.005. Medline:16831642
20. Razmjou H, Bean A, van Osnabrugge V, et al. Cross-sectional and longitudinal construct validity of two rotator cuff disease-specific outcome measures. BMC Musculoskelet Disord. 2006;13:7–26. Medline:16533405
21. Razmjou H, Bean A, Macdermid JC, et al. Convergent validity of the constant-murley outcome measure in patients with rotator cuff disease. Physiother Can. 2008;60(1):72–9. doi:10.3138/physio/60/1/72. Medline:20145743
22. Stratford PW, Kennedy DM. Does parallel item content on WOMAC's pain and function subscales limit its ability to detect change in functional status? BMC Musculoskelet Disord. 2004;5(1):17. doi:10.1186/1471-2474-5-17. Medline:15189563
23. Wessel J, Razmjou H, Mewa Y, et al. The factor validity of the Western Ontario Rotator Cuff index. BMC Musculoskelet Disord. 2005;6(1):22. doi:10.1186/1471-2474-6-22. Medline:15871742
24. Streiner DL. Clinimetrics vs. psychometrics: an unnecessary distinction. J Clin Epidemiol. 2003;56(12):1142–5, discussion 1146–9. doi:10.1016/j.jclinepi.2003.08.011. Medline:14680661
25. Fayers PM, Hand DJ, Bjordal K, et al. Causal indicators in quality of life research. Qual Life Res. 1997;6(5):393–406. doi:10.1023/A:1018491512095. Medline:9290306
26. Fayers PM, Hand DJ. Factor analysis, causal indicators and quality of life. Qual Life Res. 1997;6(2):139–50. doi:10.1023/A:1026490117121. Medline:9161114
27. Mokkink LB, Terwee CB, Knol DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. BMC Med Res Methodol. 2010;10(1):22. doi:10.1186/1471-2288-10-22. Medline:20298572
28. World Health Organization. International classification of functioning, disability and health. Geneva: The Organization; 2001.
29. Razmjou H, Davis AM, Jaglal SB, et al. Cross-sectional analysis of baseline differences of candidates for rotator cuff surgery: a sex and gender perspective. BMC Musculoskelet Disord. 2009;10(1):26. doi:10.1186/1471-2474-10-26. Medline:19239706
30. Holtby R, Razmjou H. Measurement properties of the Western Ontario rotator cuff outcome measure: a preliminary report. J Shoulder Elbow Surg. 2005;14(5):506–10. doi:10.1016/j.jse.2005.02.017. Medline:16194742
31. Michener LA, McClure PW, Sennett BJ. American Shoulder and Elbow Surgeons Standardized Shoulder Assessment Form, patient self-report section: reliability, validity, and responsiveness. J Shoulder Elbow Surg. 2002;11(6):587–94. doi:10.1067/mse.2002.127096. Medline:12469084
32. Roy JS, MacDermid JC, Woodhouse LJ. A systematic review of the psychometric properties of the Constant-Murley score. J Shoulder Elbow Surg. 2010;19(1):157–64. doi:10.1016/j.jse.2009.04.008. Medline:19559630
33. Johansson KM, Adolfsson LE. Intraobserver and interobserver reliability for the strength test in the Constant-Murley shoulder assessment. J Shoulder Elbow Surg. 2005;14(3):273–8. doi:10.1016/j.jse.2004.08.001. Medline:15889026
34. Finch E, Brooks D, Stratford P, et al. Physical rehabilitation outcome measures: a guide to enhanced clinical decision making. Hamilton (ON): Canadian Physiotherapy Association; 2002.
35. Poolman RW, Swiontkowski MF, Fairbank JC, et al. Outcome instruments: rationale for their use. J Bone Joint Surg Am. 2009;91(Suppl 3):41–9. doi:10.2106/JBJS.H.01551. Medline:19411499
36. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007;60(1):34–42. doi:10.1016/j.jclinepi.2006.03.012. Medline:17161752
37. Byrne BM. Structural equation modeling with AMOS: basic concepts, applications, and programming. Mahwah (NJ): Lawrence Erlbaum; 2001.
38. Streiner DL. Being inconsistent about consistency: when coefficient alpha does and doesn't matter. J Pers Assess. 2003;80(3):217–22. doi:10.1207/S15327752JPA8003_01. Medline:12763696