JSES International 4 (2020) 197–201

Contents lists available at ScienceDirect

JSES International

journal homepage: www.jsesinternational.org

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

Rochelle Furtado, MSc [a,b,*], Joy C. MacDermid, PT, PhD [a,b,c], Dianne M. Bryant, PhD [a,b], Kenneth J. Faber, MD, MPHE, FRCSC [b,c], George S. Athwal, MD, FRCSC [b,c]

[a] Department of Physiotherapy, School of Health and Rehabilitation Sciences, Western University, London, ON, Canada
[b] Collaborative Program in Musculoskeletal Health Research, Bone and Joint Institute, Western University, London, ON, Canada
[c] Roth McFarlane Hand and Upper Limb Centre, St Joseph's Hospital, London, ON, Canada

ARTICLE INFO

Keywords:
Rotator cuff disorders
reproducibility
agreement
reliability
shoulders
psychometrics

Level of evidence: Basic Science Study; Validation of Outcome Instrument

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 evaluated prospectively for reproducibility (reliability and agreement) and floor and ceiling effects.

Methods: Patients (N = 162) with rotator cuff disorders completed the Short-WORC at baseline. From this cohort, 47 patients underwent measurement of test-retest reliability within 2 to 7 days. We used the Cronbach α to determine internal consistency and the intraclass correlation coefficient ($ICC_{2,1}$) to assess test-retest reliability. To evaluate parameters of agreement, the standard error of measurement, minimal detectable change (based on a 90% confidence interval), and Bland-Altman plots were used.

Results: The Cronbach α was 0.82 at baseline, and the intraclass correlation coefficient ($ICC_{2,1}$) was 0.87. The agreement parameter was 8.4 for the standard error of measurement of agreement, and the limits of agreement fell within the range of −22.9 to 23.8. The Short-WORC is reliable over time and reflective of a patient's true score after an intervention.

Conclusions: The Short-WORC demonstrated strong reproducibility parameters and can be used for patients with rotator cuff disorders. The Short-WORC indicated no systematic bias and was reflective of the true score of both individual patients and groups of patients at 2 time points.

© 2019 The Author(s). Published by Elsevier Inc. on behalf of American Shoulder and Elbow Surgeons. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Rotator cuff disorders (RCDs) are a common cause of impairment and activity limitation, resulting in a loss of quality of life (QoL).[2] The prevalence of partial- and full-thickness rotator cuff tears is greater than 60% in symptomatic patients older than 60 years.[2,6] Therefore, the primary goal of both surgery and rehabilitation is to improve function and QoL in patients with RCDs.[2,21]

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.[13] Through theoretical and clinical principles supported by a factor analysis, the Western Ontario Rotator Cuff Index (WORC)[18] was reduced from 21 items to 7 items from the domains of work and lifestyle.[3,7,16] The Short-WORC consists of a smaller number of items that focus on activity limitations and generates a single summary score without the 5 domain scores generated by the original version of the WORC.[10] In 2012, Razmjou et al found strong psychometric properties for the Short-WORC and suggested that it reduces response burden.[25] Shortly thereafter, Dewan et al[9] found excellent reliability, validity, and responsiveness when extracting scores from the full WORC.[2,19] This collection of work suggests that the Short-WORC has excellent psychometric properties compared with the full WORC and other patient-reported outcome (PRO) measures.[10] However, no studies have prospectively evaluated the reproducibility (reliability and agreement) of the Short-WORC.[2,7,19]

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.[2,8,23] 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).[2,8,23] In addition, the property of agreement focuses on measurement error and evaluates the proximity of scores derived from repeated measurements. Agreement is investigated through

Ethics approval was granted by the Western University Research Ethics Board.

* Corresponding author: Rochelle Furtado, MSc, Department of Physiotherapy, School of Health and Rehabilitation Sciences, Western University, 1201 Western Road, London, ON, Canada.
E-mail address: rfurtad5@uwo.ca (R. Furtado).

https://doi.org/10.1016/j.jses.2019.10.110
2468-6026/© 2019 The Author(s). Published by Elsevier Inc. on behalf of American Shoulder and Elbow Surgeons. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

absolute reliability coefficients (standard error of measurement [SEM], minimal detectable change [MDC]) and Bland-Altman (BA) plots.[2,15]

PRO measures must demonstrate both reliability for discriminative applications and agreement to discern real change from error.[15,23] Therefore, it is critical to examine both reliability and agreement in outcome measures.[2] Thus, the purpose of this study was to evaluate the 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, Ontario, Canada.

Participants

Patients 18 years or older, and received a diagnosis of RCD were advised to undergo (surgical or rehabilitation) at the Hand and Upper Limb Centre were eligible for the study. Patients with upper-extremity fractures, adhesive capsulitis, shoulder instability, infection, tumors, and/or labral, cartilage, and ligamentous tears were excluded from the study. Among the patients (N = 162) who completed the individual items on the Short-WORC at baseline, a subset of 47 participants in stable condition were retested within 7 days to determine test-retest reliability.

We expected to obtain test-retest reliability and internal consistency (intraclass correlation coefficient [ICC]) of 0.90, as shown in previous studies.[2,12] Therefore, we calculated a sample size required to determine whether the reliability of the Short-WORC exceeds a 0.95 confidence interval (CI) around a power of 0.80.[5]

Outcome measures

The WORC was originally shortened to the 7-item Short-WORC from the domains of work and lifestyle and validated by Razmjou et al.[7] 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) and 100% (high QoL). The Short-WORC cannot be scored if items are missing.[7,10]

Statistical analysis

Data were assessed for completeness, the percentage of missing data, the presence of outliers, and floor and ceiling effects. The data set was tested for normality but was shown 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 the sample size (n) is larger than 30 participants.[23] Therefore, we used parametric statistics for our analysis as our sample size was greater than 30 participants.

SPSS software (version 24; IBM, Armonk, NY, USA) was used for data analysis, and $P < .05$ was considered statistically significant.

Floor and ceiling effects

The floor and ceiling effects were calculated as the percentage of patients whose total score fell between 0 and 10 (minimal scores) and between 90 and 100 (maximal scores). As suggested by McHorney and Ware, floor and ceiling effects are defined by using a cutoff of 15%.[22] Therefore, floor and ceiling effects were considered to exist if more than 15% of participants reported minimal or maximal total scores.

Reliability

Internal consistency (using the ICC) is defined as the degree to which items on a questionnaire are correlated with each other when assessed at 1 point in time.[1] The[1,14] Cronbach α was calculated with a 95% CI to assess internal consistency at baseline and at 3 months' follow-up. An α of 0.70 to 0.90 was deemed to indicate excellent internal consistency.

Test-retest reliability (longitudinal reliability) measures the extent to which consistent results are obtained at test and retest occasions in subjects in stable condition.[10] A value of 0.70 to 0.80 is deemed appropriate for comparison in research, and a value over 0.90 is deemed appropriate for clinical interpretation.[1,19,20] Test-retest scores were analyzed using a 2-way mixed model with absolute agreement to produce an intraclass correlation ($ICC_{2,1}$) with a 95% CI for a single measure.[15,26] An ICC of 0.80 was considered the minimum standard for good reliability in this study.[2,15]

Statistical hypothesis

We expected that the Short-WORC would demonstrate excellent internal consistency (Cronbach α) and test-retest reliability ($ICC_{2,1}$), defined as values of 0.80 or greater and 0.90 or greater, respectively.

Agreement parameters

Absolute reliability was assessed by calculating the following statistics: SEM and minimal detectable change based on a 90% confidence interval (MDC_{90}). The SEM of agreement ($SEM_{agreement}$) was calculated using the following equation[7]:

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

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

Assuming that our data verify the 2 required assumptions for estimation of MDC_{90} (ie, normally distributed data and no systematic bias), we used the SEM to calculate the MDC_{90} with the following equation[2,9]:

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

The SEM provides the estimate of measurement error in the same units as the original measurement, and MDC_{90} is the minimum amount of change required to be 90% confident that a change has occurred over a period without measurement error.[2] The 95% CI for MDC_{90} was calculated as d ± $MDC_{90,}$ in which d is the mean difference.[2,9]

To calculate the real change over time between groups of patients, we calculated the minimal detectable change for a group given a 90% confidence (MDC90) using the following formula: $MDC_{group} = MDC_{90}/\sqrt{n} \times 1.64$, in which n is the sample size of the group.[7,27] Smaller SEM and MDC values indicate smaller measurement errors.[2]

Table I
Patient baseline characteristics (N = 162)

Variable	Data
Age, mean ± SD, yr	61.2 ± 16.3 (162 [100%])
Sex, n (%)	
Male	79 (48.8)
Female	90 (55.5)
Affected shoulder, n (%)	
Left	54 (33.3)
Right	94 (58.1)
Bilateral	14 (8.6)
Occupational information, n (%)	
Employed	80 (49.4)
LOA	12 (7.4)
Unemployed	10 (6.2)
Retired	60 (37.1)

SD, standard deviation.

BA plots

BA plots were used for plotting the limits of agreement (LOAs). The LOA is the difference between scores at time 1 and time 2 of the test-retest period against the mean score for the 2 points. The BA plots produce an image of the results that can be used to evaluate systematic variability (bias), the presence of outliers, and homoscedasticity.[3,4,6]

Results

The demographic characteristics of the study population are presented in Table I.

Reliability

There were no floor or ceiling effects; therefore, the Short-WORC is appropriate for rotator cuff pathology. Internal consistency (Cronbach α [95% CI]) was excellent at the baseline assessment, with a value of 0.82 (N = 162), and at 3 months postoperatively, with a value of 0.87 (n = 51). Therefore, the Short-WORC is stable and reflects the true score of the patient. Test-retest reliability was excellent ($ICC_{2,1}$ = 0.87) (Table II).

Agreement parameters

Reported values for $SEM_{agreement}$ (8.4), MDC_{90} (19.5), and $MDC_{90group}$ (1.7) are shown in Table III. These findings indicate that the Short-WORC is an excellent measure of change within a group of patients.

BA plots

The 95% LOAs for test-retest scores are presented in Table II. Visual inspection shows the random scatter of most points to be within the 95% LOAs and represents negligible systematic bias (error) between scores for the Short-WORC (Fig. 1). Short-WORC scores are stable between time intervals and are reflective of a patient's true score of an intervention.

Table II
Longitudinal reliability of Short-WORC

	Test-retest reliability
Test mean (SD)	45.6 (23.9)
Retest mean (SD)	45.1 (23.1)
d (SD)	0.5 (11.9)
95% CI for d	−2.8 to 3.4
95% LOAs	−22.9 to 23.8
ICC (95% CI)	0.87 (0.79-0.92)

Short-WORC, short version of Western Ontario Rotator Cuff Index; *SD*, standard deviation; *d*, mean difference; *CI*, confidence interval; LOAs, limits of agreement; *ICC*, intraclass correlation coefficient.

Table III
Reproducibility: agreement parameters of Short-WORC

	Value
$SEM_{agreement}$ (95% CI)	8.4
$MDC_{90individual}$	19.5 (−19 to 20)
$MDC_{90group}$	1.7

Short-WORC, short version of Western Ontario Rotator Cuff Index; $SEM_{agreement}$, standard error of measurement of agreement; *CI*, confidence interval; MDC_{90}, the minimal detectable change for a group given a 90% confidence.

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 findings of previous studies that assessed these properties retrospectively.[2,7] 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[2,7,19] when making decisions about patient QoL and changes in QoL.[24,28]

In this study, we did not observe floor or ceiling effects, which is also consistent with previously published work,[2,7,19] suggesting that the Short-WORC is well suited to detect both improvement and worsening in the RCD population. The internal consistency (0.82) was acceptable and similar to that reported by Razmjou et al (0.89) and Dewan et al[9] (0.84) at baseline[2,7] and was comparable to the Cronbach α (0.85-0.92) of the original WORC depending on the translation.[10,11,17] Because it has been suggested that values exceeding 0.90 indicate redundancy, the Short-WORC may be more efficient than the WORC.[2,7]

An ICC of 0.90 or greater can be difficult to obtain; however, previous literature considered a measure reliable if the point estimate exceeded 0.75.[1,2,19] The ICCs found in our study were similar to those of previously published work and the WORC.[7,10,11,17] The $ICC_{2,1}$ value (0.87) exceeded the benchmark of 0.75. Thus, our study provides strong evidence that the Short-WORC has excellent reliability across multiple contexts. On the basis of our narrow CI, we can be confident that our estimate is precise and exceeds minimum expectations.

The $SEM_{agreement}$ value of 8.4 for the Short-WORC found in this study indicates that there is a 68% chance (1 ± SEM) that the true score on the Short-WORC for an individual assessed at a single point in time lies within 8.4 points of the measured score. We used the $ICC_{2,1}$ absolute agreement value to calculate the SEM instead of the Cronbach α and did not choose to use the Cronbach α to estimate the SEM. Instead, we used the $SEM_{agreement}$ value to compute MDC,[1,2,15] as it expresses the measurement error through the systematic difference between test and retest scores, which is otherwise ignored with the SEM consistency value.[15]

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),[2] we observed that the MDC value is higher for the Short-WORC (19.5). This could be a result of fewer items on the Short-WORC, therefore producing greater variability.

The low MDC_{group} value indicates that the Short-WORC is an excellent measure of change within a group of patients. When both MDC_{90} values (individual vs. group) are compared, the

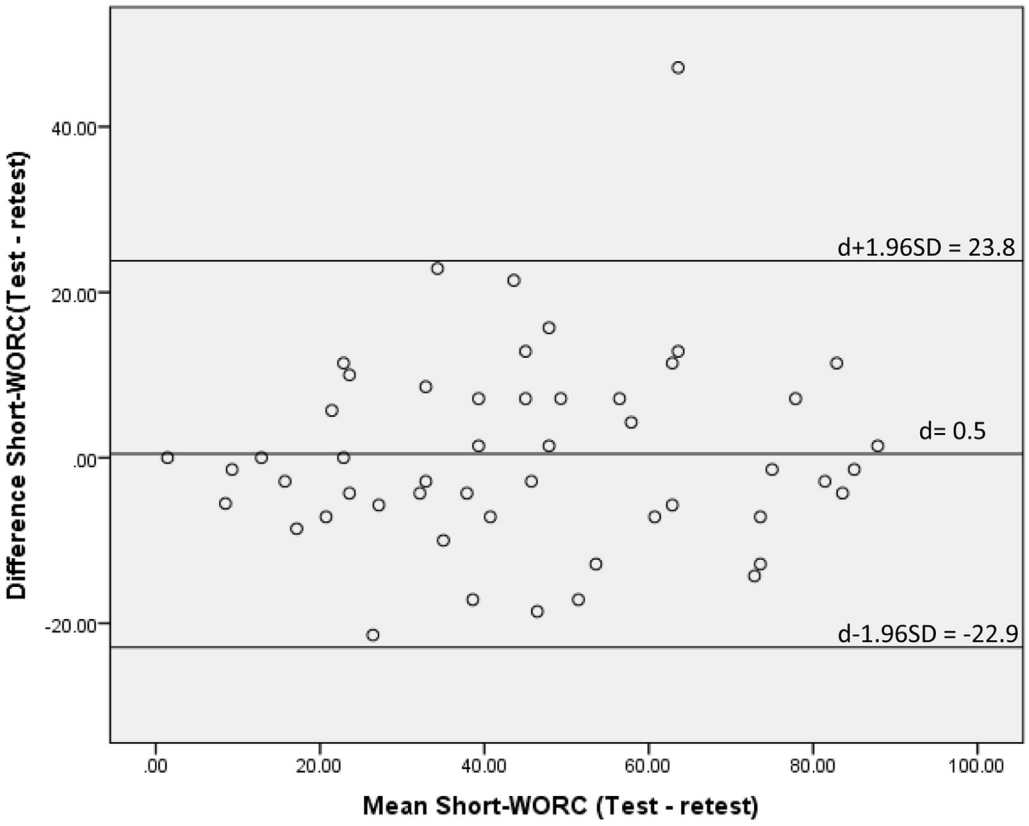

Figure 1 Bland-Altman limits-of-agreement plot between test and retest scores on short version of Western Ontario Rotator Cuff Index *(Short-WORC)* (n = 51). The *middle line* represents the mean of the individual differences *(d)*, and *upper and lower lines* represent the 95% limits of agreement. Differences lie between d ± 1.96 standard deviation *(SD)* of the mean difference.

Short-WORC is better at measuring change for a group of patients than for an individual patient.[27] As shown in the literature, a smaller value for $MDC_{90group}$ than for $MDC_{90individual}$ aligns with agreement parameters reported for other PRO measures. This is an expected finding because the formula for $MDC_{90group}$ is dependent on the square root of the sample size, unlike that for $MDC_{90individual}$, which is dependent on the square root of 2 and the error band around the mean difference of 2 measurements. This is further evident as the group effect will always average out any differences that would normally be highlighted in the individual effect. Therefore, the variability will always be higher for $MDC_{90individual}$ compared with $MDC_{90group}$. However, measuring both groups of patients and individual patients is important to ensure that the measure is reliable when assessing an individual patient over a test-retest interval and over a period between groups of patients after an intervention.[2,23]

The LOAs 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% LOAs (−22.9 to 23.8) reported in our study reflect large within-individual variability and hence limited usefulness of measures for individual comparisons. We used the retest assessment period of 2 to 7 days as a stable period for patients because it is long enough to prevent recall bias but short enough to expect that no clinical change has occurred given that RCD is a chronic condition. This interval was sufficient according to previous literature but can allow some potential for circumstances to destabilize a patient's condition.[2] Our assumption of considering 1 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 this study suggest that the Short-WORC can replace the 21-item WORC for both clinical and research applications. Although we found high values for the LOAs (−22.9 to 23.8), they are similar to those of the WORC (−22.7 to 20.1) and those in our previously published work (−26.5 to 22.3).[2] The agreement parameters are also in accordance with our previously published work.[2,7,19]

Overall, our findings are consistent with values obtained when the Short-WORC was extracted from its full parent version. Lower internal consistency and reliability with wider variations between test-retest scores can be expected when using abbreviated questionnaires.[23] The goal of shortening a questionnaire is to reduce patient and/or administrative burden while retaining the conceptual linkage to the intended construct and sufficient psychometric equivalence.[24,28] We assumed that patients required less time to complete the 7 items of the Short-WORC than the 21 items of the original WORC, although we did not directly measure time. 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 assessments and should include longitudinal evaluations of responsiveness. Although our previous work[9] supported the responsiveness of the Short-WORC, it was conducted with data collected using 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 is consistent. In addition, all of the studies to date have been conducted at specialty shoulder surgery clinics; therefore, assessment of populations in

different contexts or that include a broader spectrum of RCDs would clarify whether these measurement properties exist in multiple contexts of the disorder.

Conclusion

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

Disclaimer

Joy C. MacDermid was supported by a Canadian Institute for Health Research Chair in Gender, Work and Health and the Dr. James Roth Research Chair in Musculoskeletal Measurement and Knowledge Translation.

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

References

1. Beaton DE, Katz JN, Fossel AH, Wright JG, Tarasuk V, Bombardier C. 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;14:128–42.
2. 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;10:571–8.
3. Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res 1999;8:135–60.
4. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307–10.
5. Bonett DG. Sample size requirements for estimating intraclass correlations with desired precision. Stat Med 2002;21:1331–5. https://doi.org/10.1002/sim.1108.
6. Brennan P, Silman A. Statistical methods for assessing observer variability in clinical measures. BMJ 1992;304:1491–4.
7. Chen HM, Hsieh CL, Sing Kai L, Liaw LJ, Chen SM, Lin JH. The test-retest reliability of 2 mobility performance tests in patients with chronic stroke. Neurorehabil Neural Repair 2007;21:347–52. https://doi.org/10.1177/1545968306297864.
8. de Vet HC, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. J Clin Epidemiol 2006;59:1033–9. https://doi.org/10.1016/j.jclinepi.2005.10.015.
9. 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;29:281–91. https://doi.org/10.1016/j.jht.2015.11.007.
10. 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;27:758–64.
11. El O, Bircan C, Gulbahar S, Demiral Y, Sahin E, Baydar M, et al. The reliability and validity of the Turkish version of the Western Ontario Rotator Cuff Index. Rheumatol Int 2006;26:1101–8. https://doi.org/10.1007/s00296-006-0151-2.
12. Folwer JJP, Chevannes M. Practical statistics for nursing and health care. West Sussex, UK: John Wiley & Sons; 2002.
13. Furtado R, MacDermid JC. Clinimetrics: Short Western Ontario Rotator Cuff Index. J Physiother 2019;65:56. https://doi.org/10.1016/j.jphys.2018.10.005.
14. Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. J Chronic Dis 1987;40:171–8.
15. Harvill LM. Standard error of measurement. Educ Meas Issues Pract 1991;10:33–41.
16. Holtby R, Razmjou H. Measurement properties of the Western Ontario rotator cuff outcome measure: a preliminary report. J Shoulder Elbow Surg 2005;14:506–10. https://doi.org/10.1016/j.jse.2005.02.017.
17. Kawabata M, Miyata T, Tatsuki H, Nakai D, Sato M, Kashiwazaki Y, et al. Reproducibility and validity of the Japanese version of the Western Ontario Rotator Cuff Index. J Orthop Sci 2013;18:705–11. https://doi.org/10.1007/s00776-013-0426-x.
18. 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;13:84–92. https://doi.org/10.1097/00042752-200303000-00004.
19. Kottner J, Audige L, Brorson S, Donner A, Gajewski BJ, Hrojartsson A, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. J Clin Epidemiol 2011;64:96–106. https://doi.org/10.1016/j.ijnurstu.2011.01.016.
20. 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. https://doi.org/10.2519/jospt.2015.5033.
21. 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;13:593–8. https://doi.org/10.1016/j.jse.2004.03.009.
22. McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? Qual Life Res 1995;4:293–307.
23. Norman GR, Streiner DL. The normal distribution. In: Biostatistics: the bare essentials. Hamilton, Ontario: B.C. Decker; 2008. p. 31–6.
24. Portney LG, Watkins MP. Foundations of clinical research: applications to practice. Upper Saddle River, NJ: Prentice Hall Health; 2009.
25. Razmjou H, Stratford P, Holtby R. A shortened version of the Western Ontario Rotator Cuff Disability Index: development and measurement properties. Physiother Can 2012;64:135–44. https://doi.org/10.3138/ptc.2010-51.
26. Schuck P. Assessing reproducibility for interval data in health-related quality of life questionnaires: which coefficient should be used? Qual Life Res 2004;13:571–86. https://doi.org/10.1023/B:QURE.0000021318.92272.2a.
27. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979;86:420–8.
28. Streiner DL, Norman GR. Health measurement scales: a practical guide to their development and use. 4th ed. Oxford: Oxford University Press; 2008.