

Cross-cultural adaptation and validation of the Chinese version of the shoulder pain and disability index in patients with symptomatic shoulder pain

A prospective case series

Wei Wang, MD^a, Zhen-yu Jia, MD^b, Jiao Liu, MD^c, Qing-yun Xie, MD^a, Jin Cui, MD^b, Wei Zheng, MD^a, Wei-dong Xu, MD^{b,*}

Abstract

The aim of this study is to cross-culturally adapt and translate the original version of SPADI into Chinese (C-SPADI), and to test the reliability, validity, and responsiveness of the C-SPADI.

This research was a test of previously developed diagnostic criteria in a series of consecutive patients with universally applied gold standard. The original version was translated into Chinese according to international recognized standards. Patients who were diagnosed with a shoulder disorder and underwent shoulder arthroscopic treatments from 2014 to 2015 were enrolled in our study. Each participant was asked to finish the C-SPADI, the Short-Form 36 (SF-36), and the Oxford Shoulder score (OSS) at first visit. The C-SPADI was completed a second time with an interval of 7 days. Six months after arthroscopic treatments, the C-SPADI was completed a third time for responsiveness evaluation. The Cronbach alpha, intraclass correlation coefficient (ICC), standard error of measurement (SEM), minimally detectable change (MDC), Pearson correlation coefficient (r), effect size (ES), and standardized response mean (SRM) were calculated to evaluate the reliability, validity, and responsiveness of C-SPADI respectively.

The original version of the SPADI was well adapted and translated into Chinese. The Cronbach alpha ranged from 0.812 to 0.912 in all subscales and total scale of the C-SPADI, indicating good or excellent internal consistency. The test-retest reliability (ICC = 0.887–0.915, SEM = 5.47, MDC = 15.16) was proved to be good or excellent. Moderate or good correlations ($r = 0.556–0.672$) were obtained between the C-SPADI and the OSS, physical subscales of SF-36; and poor, fair, or moderate correlations ($r = 0.038–0.492$) were obtained between the C-SPADI and mental subscales of SF-36, which, adequately illustrated good discriminant validity in the C-SPADI. Additionally, the responsiveness was considered good in the C-SPADI (SRM = 1.58–2.44, ES = 1.79–2.17).

The C-SPADI was documented to be a reliable, valid, and responsible instrument for self-assessment of patients with shoulder disorders in China.

Level of Evidence: Level II.

Abbreviations: AAOS = the American Association of Orthopaedic Surgeons, BP = bodily pain, CMSOS = the Constant–Murley shoulder outcome score, DASH = Disability of the Arm, Shoulder and Hand Questionnaire, ES = effect size, GH = general health, HRQOL = health-related quality of life, ICC = intraclass correlation coefficient, MDC = minimally detectable change, MH = mental health, OSS = the Oxford Shoulder score, PF = physical functioning, QoL = quality of life, RE = role-emotional, RP = role-physical, SEM = standard error of measurement, SF = social functioning, SF-36 = the Short-Form 36, SPADI = the Shoulder Pain and Disability Index, SRM = standardized response mean, VT = vitality.

Keywords: cross-cultural adaptation, reliability, responsiveness, SPADI, validation, validity

Editor: Leonardo Roever.

WW, Z-yJ, JL, and Q-yX contributed equally to this work as co-first authors.

The authors have no conflicts of interest to disclose.

^a Department of Orthopedics, Chengdu Military General Hospital, Chengdu city,

^b Department of Orthopedics, Changhai Hospital, ^c Department of Hepatobiliary Surgery, Shanghai Public Health Clinical Center Affiliated to Fudan University, Shanghai, China.

* Correspondence: Wei-dong Xu, Department of Orthopedic, Changhai Hospital, Second Military Medical University, No 168, Changhai Road, Shanghai 200433, China (e-mail: wdxushanghai@sina.com).

Copyright © 2018 the Author(s). Published by Wolters Kluwer Health, Inc. This is an open access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

Medicine (2018) 97:26(e11227)

Received: 16 February 2017 / Accepted: 21 May 2018

<http://dx.doi.org/10.1097/MD.00000000000011227>

1. Introduction

Shoulder pain is the third most common musculoskeletal pain with lifetime prevalence in the general population of 6.7% to 66.7%.^[1] Although shoulder pain is not a life-threatening disease, it is still a major problem in China because pain and stiffness in the shoulder may lead to labor disability, work absenteeism, and large medical costs. As reported, the cross-sectional prevalence of neck and shoulder pain is 41.1% in Chinese adolescents.^[2]

Since the 1980s, the evaluation of health-related quality of life (HRQOL) was developed to specifically ascertain the patients' needs and problems of everyday life and to enable patients themselves provide important information about their physical and psychological status.^[3] Disease-specific scales are used to assess specific patient groups, such as the most commonly used instruments for shoulder disorders: Shoulder Pain and Disability Index (SPADI), Disability of the Arm,

Shoulder and Hand Questionnaire (DASH), and the Oxford Shoulder Score (OSS).^[4–6]

Because of the increasing number of large multicenter international studies and the requirement for globally meaningful epidemiologic and/or therapeutic study results, there is a need for cross-cultural adaptation and validation of HRQOL scales.^[7] To avoid the evaluation error caused by cultural differences, it is necessary to test the psychometric properties of the scale rather than simply translating the content.^[8,9] The SPADI is a simple disease-specific scale to evaluate the impact of shoulder disorders on the general quality of life.^[4] It has been well adapted in 12 languages and widely proved to be a valid and reliable instrument.^[4,10–20] However, instruments for shoulder are still in shortage in China. Now there is no one Chinese version of SPADI even though China has the largest population of patients with shoulder disorders. As applied in this study, psychometric properties including reliability, validity, and responsiveness were tested for validation according to the advice and guidelines of the Consensus-based Standards for the Selection of Health Status Measurement Instruments study.^[21]

Therefore, the aim of this study was to translate and adapt the SPADI into a Chinese version (C-SPADI) and evaluated the reliability, validity, and responsiveness of the C-SPADI in a cohort of native Chinese-speaking patients with shoulder disorders.

2. Patients and methods

2.1. Translation and cross-cultural adaptation

The procedure of translation and cross-cultural adaptation was performed following the guidelines of the American Association of Orthopaedic Surgeons (AAOS) Outcome Committee.^[22] There are 5 steps needed to perform.

Step 1—Forward translation: respective translation by 2 bilingual translators whose first language was Chinese (a professor of our department and an English professor);

Step 2—Synthesis of 2 translations: a discussion was held to integrate the 2 independent forward translation drafts, later we obtained the primary Chinese version of SPADI;

Step 3—Back translation: back-translated into English by 2 bilingual translators whose first language was English, and the 2 translators have no knowledge about the original SPADI;

Step 4—Creation of prefinal version: a prefinal version of C-SPADI created by an expert committee after the discrepancies, ambiguities, and other language expression issues existed in the Chinese version were solved;

Step 5—Test of prefinal version: 20 patients with shoulder disorders completed the prefinal C-SPADI for assessment, and feedbacks were collected. A discussion was held for final adjustments according to these feedbacks, and the final version of C-SPADI was obtained.

2.2. Participants and data collection

This research was a test of previously developed diagnostic criteria in a series of consecutive patients with universally applied gold standard. Patients enrolled in this study were mainly consecutively recruited from those with symptomatic shoulder disorders admitted in Shanghai Changhai Hospital and Chengdu Military General Hospital between March 2014 and March 2015. Inclusion criteria were as follows: patients with shoulder disorders and scheduled for shoulder arthroscopic treatment; able to read Chinese and fill in the scales independently. Patients

who were under 18 years old and had other disorders that could affect shoulder were excluded. Our study met the quality criteria proposed by Terwee et al^[23] for measurement properties of health status scales, which required the results from at least 100 patients to perform internal consistency analysis and from at least 50 patients for floor or ceiling effects, reliability, and validity analyses. Informed written consent was obtained from each participant. The study was approved by the clinical research ethics committee of our hospital (No. CHEC 2014-071, Date: 2014/01/10).

All patients were asked to provide demographic information on the first day of enrollment, and to independently complete 3 scales included the C-SPADI, OSS, and SF-36 in a quiet meeting room. One week after the first filling-in, also the day before the arthroscopic surgery, they completed the C-SPADI for the second time to evaluate the test–retest reliability of the scale. Six months after the surgery, when the patients came to our hospital for regularly check, they completed the C-SPADI for the third time to help evaluate the responsiveness.

2.3. Scales

The SPADI was designed to measure the impact of shoulder pathology in terms of pain and disability.^[4] It is a self-administered questionnaire including pain subscale (5 items) and disability subscale (8 items). Each item is rated on a visual analogue scale from 0 to 11, with 0 representing no pain/no difficulty and 11 indicating worst pain/most difficulty. The score of total scale consists of the score of 2 subscales and all scores are converted to a 0 to 100 scale.

The SF-36 was a widely used instrument that measured 8 subscales of health including physical functioning (PF), role-physical (RP), bodily pain (BP), general health (GH), vitality (VT), social functioning (SF), role-emotional (RE), and mental health (MH). Each subscale score was transformed into a 0 to 100 scale; a lower score indicated worse quality of life and severer disability.^[24] The OSS was a shoulder pain-specific questionnaire for the evaluation of pain perception and daily function in patients suffering from shoulder pain using 12 items. Each item scored from 0 to 4 and the total score ranged from 0 (worst) to 48 (best). A lower score indicated a worse status of shoulder joint.^[6] Chinese versions of the above 2 scales were available, which had been proved reliable, valid, and responsive.^[25,26]

2.4. Psychometric assessments and statistical analysis

Statistical package for the Social Sciences, version 20.0 (SPSS, Chicago, IL) was used for statistical analysis. *P* value of .05 or less was considered statistically significant.

Test–retest reliability was evaluated to assess the reliability of C-SPADI using intraclass correlation coefficient (ICC) and the Bland–Altman plot. The outcomes were classified into 5 grades ($r = 0.81–1.0$ excellent; $0.61–0.80$ very good; $0.41–0.60$ good; $0.21–0.40$ fair; and $0.00–0.20$ poor).^[27] The Bland–Altman plot could be used to measure within-subject variation and limits of agreement.^[28] Measurement error was assessed by calculating the standard error of measurement (SEM) and the minimally detectable change (MDC).^[29] SEM was calculated as the square root of the within-subject total variance of an ANOVA analysis. MDC_{95} was calculated as $1.96 \times \sqrt{2} \times SEM$.^[30] Internal consistency was assessed using Cronbach alpha. The value of Cronbach alpha between 0.70 and 0.95 indicated a good internal consistency.^[23]

Table 1**Demographic and clinical characteristics of participants.**

Characteristics	Number or mean \pm SD
Age	51.3 \pm 11.2
Range	18–73
Gender	
Female	48 (38.1%)
male	78 (61.9%)
Affected side	
Right	65 (51.6%)
Left	61 (48.4%)
Dominant side	
Dominant	74 (58.7%)
Nondominant	52 (41.3%)
BMI	23.3 \pm 4.5
Diagnosis	
Rotator cuff tear	62 (49.2%)
Frozen shoulder	39 (31.0%)
LHBT injury	11 (8.7%)
SLAP injury	9 (7.1%)
Bankart injury	5 (4.0%)

BMI=body mass index, LHBT=long head of biceps tendon injury, SLAP=superior labrum from anterior to posterior injury.

The validity of the C-SPADI was assessed by evaluating both of the content validity and construct validity. The content validity consists of the assessments of comprehensiveness and relevance of items.^[31] The item response rate, ceiling/floor effects, and patient feedback were the 3 indexes for comprehensiveness assessment. If the response rate for each item in the scale was $>95\%$, the ceiling/floor effect of each subscale was $<15\%$, and no difficulties in understanding the items were fed back from the participants, the scale was considered to have good comprehensiveness.^[23] In addition, a rehabilitation medicine expert and 3 orthopedics specialists were invited to help us to judge whether the items were relevant for the construct to be measured and for the patients with symptomatic shoulder disorders population. Because there is no gold standard for evaluating the validity of C-SPADI, the hypotheses testing method was employed to evaluate the construct validity.^[31] In this study, we selected the OSS and SF-36 as the control scales for the C-SPADI. Based on the content of each scale, we hypothesized that the subscales and overall scale of C-SPADI should be well correlated with the OSS and the physical subscales of SF-36 (PF, RP, BP, GH), but poorly with the mental subscales of SF-36 (VT, SF, RE, MH). In addition, because the OSS is specifically designed to assess the patient's function in shoulder area, and the SF-36 is only a generic scale, we hypothesized that the correlation between the C-SPADI and the OSS should be better than that with any subscales of SF-36. Based on the above hypotheses, the Pearson correlation coefficient (r) of the C-SPADI with the OSS and SF-36 were calculated.

Table 2**Test-retest reliability and responsiveness of the C-SPADI.**

Scale	1st test (mean \pm SD)*	2nd test (mean \pm SD)*	3rd test (mean \pm SD)*	ICC (CI range)	SEM	MDC ₉₅	ES	SRM
Overall scale	60.0 \pm 13.3	59.6 \pm 15.0	42.5 \pm 9.5	0.900 (0.861–0.929)	4.17	11.56	–2.02	–2.44
Pain subscale	70.5 \pm 12.1	70.1 \pm 13.8	54.8 \pm 12.0	0.887 (0.844–0.919)	4.05	11.23	–1.79	–1.58
Disability subscale	49.5 \pm 14.9	49.1 \pm 16.6	30.2 \pm 9.9	0.915 (0.881–0.939)	5.47	15.16	–2.17	–2.29

CI=95% confidence interval, C-SPADI=Chinese version of Shoulder Pain and Disability Index, ES=effect size, ICC=intraclass correlation coefficient, MDC=minimally detectable change, SD=standard deviation, SEM=standard error of measurement, SRM=standardized response mean.

*The 1st test was conducted at the beginning of this research (126 patients), the 2nd test was conducted 1 week later to calculate the test-retest reliability (ICC) of the C-SPADI (126 patients), and the 3rd test was conducted 6 months later to calculate the responsiveness (ES, SRM) of the C-SPADI (102 patients).

Correlations were categorized as follows: poor ($r=0-0.20$), fair ($r=0.21-0.40$), moderate ($r=0.41-0.60$), very good ($r=0.61-0.80$), or excellent ($r=0.8-1.0$).^[32]

The standardized response mean (SRM) and the effect size (ES) were calculated to assess the responsiveness by comparing the results of 1st and 3rd tests of C-SPADI. SRM was defined as the mean change between these time points divided by the SD of this change. ES was defined as the mean change between preoperative results and 6-month postoperative results divided by the SD of the preoperative C-SPADI score.^[33] ES and SRM were considered large if >0.80 , moderate between 0.51 and 0.80 and small if lower than 0.50.^[34]

3. Results

3.1. Cross-cultural adaptation

During forward and back-translation of SPADI, there were no major problems or large language difficulties existed. Small revision was made in disability subscale to adapt to the cultural background. Because the Chinese do not use “pounds” as weight units, we changed the “10 pounds” in item 7 into “10 jin.” The formula for unit conversion was as follows. No patient raised the problem of understanding in our pretest and formal study.

1 pound= 0.907 jin.

3.2. Patients

Altogether 126 (48 female/78 male) patients with shoulder disorders were recruited in this study. All of them completed the first 2 rounds scales. However, 24 patients failed to finish the scale for the third time (6 months after their shoulder arthroscopic surgery). As a result, the results of 126 subjects were included in C-SPADI's reliability and validity tests, while only 102 subjects' results were assessed in the scale's responsiveness test. Patients' detailed demographic information was listed in Table 1.

3.3. Reliability

The ICCs for the subscales and overall scale of C-SPADI were all more than 0.80 (0.887–0.915) as shown in Table 2, which indicated excellent test-retest reliability. The Bland-Altman plots revealed that the test-retest differences for the C-SPADI were centered around zero (Fig. 1), which indicated there was no systemic error in the data obtained from the first 2 rounds of completed scales and confirmed the good test-retest agreement of C-SPADI. The SEM was 5.47 on the basis of repeated measurements for test-retest. The MDC on the basis of the SEM for test-retest was 15.16 (Table 2). The Cronbach alpha of all subscales and overall scale was from 0.826 to 0.912 (Table 3), which proved the internal consistency of C-SPADI was good.

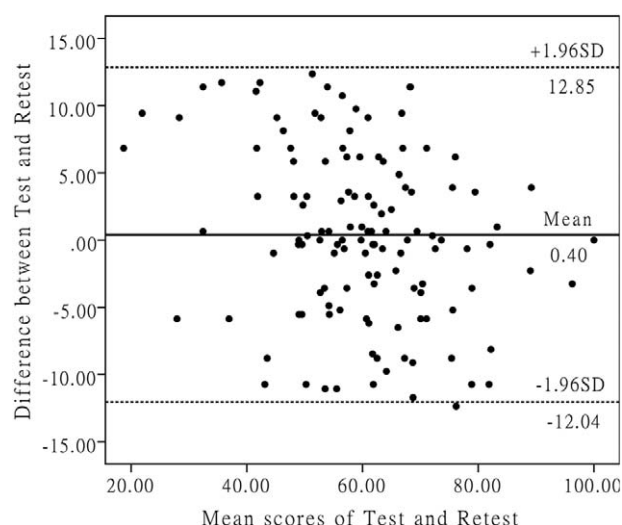


Figure 1. These are Bland–Altman plots of test–retest reliability of the C-SPADI. Each data point indicates how the difference between the 2 test sessions for an individual patient compares to the mean of the 2 sessions for scores of each C-SPADI. The interval of 2 sessions was 1 week. The dashed line shows the 95% (± 1.96 SD) limits of agreement. C-SPADI=the Chinese Version of the Shoulder Pain and Disability Index.

3.4. Validity

The score distribution showed that there was no ceiling effect (0.8%) or floor effect (0%) in overall C-SPADI scale and its subscales. No doubt and misunderstanding were reported. After the invited experts’ assessment, the data acquired from C-SPADI was adequate to evaluate the HRQOL of patients with shoulder disorders. In summary, we concluded C-SPADI had a good content validity.

Construct validity assessment data are listed in Table 4. The subscales and overall scale of C-SPADI were moderate or good correlated with the OSS and physical subscales of SF-36 ($r=0.556-0.680$), and poor, fair, or moderate correlated with the mental subscales of SF-36 ($r=0.038-0.492$). In addition, the subscales and overall scale of C-SPADI were stronger correlated with OSS ($r=0.663-0.680$) than that with subscales of SF-36 ($r=0.038-0.613$). The above results were consistent with our hypothesis, indicating a good construct validity of C-SPADI.

3.5. Responsiveness

The average scores of the overall scale and subscales of C-SPADI had all decreased after the arthroscopic treatment. And the values of ES and SRM were ranged from 1.79 to 2.17 and 1.58 to 2.44, respectively (Table 2). The results indicated good responsiveness of C-SPADI.

Table 4
Construct validity of the C-SPADI.

Correlation coefficient <i>r</i> (<i>P</i> value)*	SC-SPADI subscales		Overall score
	Pain	Disability	
SF-36 subscales			
Physical function	-0.604 (<.001)	-0.605 (<.001)	-0.613 (<.001)
Role-physical	-0.629 (<.001)	-0.670 (<.001)	-0.661 (<.001)
Bodily pain	-0.634 (<.001)	-0.672 (<.001)	-0.664 (<.001)
General health	-0.556 (<.001)	-0.617 (<.001)	-0.597 (<.001)
Vitality	0.038 (.673)	0.089 (.319)	0.067 (.455)
Social function	-0.492 (<.001)	-0.451 (<.001)	-0.476 (<.001)
Role-emotional	-0.107 (.235)	-0.010 (.908)	-0.054 (.545)
Mental health	-0.170 (.057)	-0.097 (.278)	-0.132 (.142)
OSS	-0.663 (<.001)	-0.677 (<.001)	0.680 (<.001)

C-SPADI=Chinese version of Shoulder Pain and Disability Index, OSS=Oxford shoulder score, SF-36=Short-Form 36.

* Calculated by the Pearson correlation coefficient (*r*) of the C-SPADI with OSS and SF-36.

4. Discussion

The most important finding of this study was that the C-SPADI was demonstrated to have a good score distribution, high internal consistency, excellent test–retest reliability, and notable content and construct validity. Several instruments were designed to evaluate quality of life (QoL) for patients with shoulder disorders and the most commonly used were DASH, OSS, SPADI, and the Constant–Murley shoulder outcome score (CMSOS).^[35] The OSS is by far the only shoulder scale that has been translated and adapted into Chinese, and the scales for shoulder are still in shortage in China. DASH was used to evaluate patients with upper extremity musculoskeletal conditions but not a shoulder-specific instrument. CMSOS could be affected by surgeon bias and 2 components of CMSOS were tested by the surgeon.^[36] SPADI is a widely used scale to evaluate the functioning status of patients with shoulder disorders, and it has been translated into multiple languages and its reliability, validity, and responsiveness have been well established.^[10-20] It is meaningful to translate SPADI into Chinese to serve the largest number of population in the world.

The value of ICC was ranged from 0.887 to 0.915 indicating excellent reproducibility of C-SPADI, which was similar to values of German, Slovene, Brazilian, Persian, Thai, and Arabic versions (ICC=0.84–0.95).^[10,11,17,18,20] The values of MDC and SEM were similar as previous studies.^[10,17] However, compared with the original version, the ICCs of the subscales and overall scale in C-SPADI were higher than that in the original version. It might be some real changes occurred in the original study over a 48-hour period and most of the patients (91%) were placed on analgesic medication and 50% of them received a periarticular corticosteroid injection to decrease pain and inflammation,^[4] while all patients finished the C-SPADI for the second time before any

Table 3
The internal consistency and score distribution of the C-SPADI.

Scale	No. of items	Mean \pm SD	Observed range	Cronbach alpha	Floor effect (%)*	Ceiling effect (%)*
Overall scale	13	60.0 \pm 13.3	22.1–100.0	0.912	0	0.8
Pain subscale	5	70.5 \pm 12.1	38.0–100.0	0.888	0	0.8
Disability subscale	8	49.5 \pm 14.9	6.3–100.0	0.826	0	0.8

C-SPADI=Chinese version of Shoulder Pain and Disability Index, SD=standard deviation.

* Percentage of patients with the worst (floor effect) and the best (ceiling effect) condition.

treatments in our study. The Cronbach alpha was 0.912 indicating excellent internal consistency of C-SPADI and the value was close to other language versions proving SPADI was stable in different cultures. However, the Cronbach alpha was higher for the total score than 2 subscales, which indicated the 2 subscales do not reflect separate dimensions clearly.

No ceiling effect and floor effect was observed in the overall scale of C-SPADI and all the subscales. Experts assessments also confirmed that the items of C-SPADI are good relevant for the construct to be measured and for the patients with symptomatic shoulder disorders. And there was no one item not answered in the formal research. Integrating these results, we considered that the C-SPADI has good content validity.

Correlations between the C-SPADI and the subscales of SF-36 and the OSS were generally consistent with our hypotheses, suggesting that it has good construct validity, and these results also accorded with relevant conclusions from other studies.^[10,17,18] The correlation between the C-SPADI and the OSS is the strongest, despite that the OSS is not specifically developed for patients with symptomatic shoulder disorders. But the OSS has also focused on the region of shoulder, just as the SPADI does. Although the physical subscales of SF-36 were strongly associated with the C-SPADI, it was still lower than that between the C-SPADI and the OSS. This is because that the accuracy SF-36, as a generic scale, in the function status assessment of specific types of patients is lower than that of other specific scales.^[37] Furthermore, correlations between the mental subscales of SF-36 and C-SPADI were poor, and this result was logical and consistent with that of other studies.^[10,18]

Scale's responsiveness is an important factor in determining whether it can be used in prospective clinical research. Our study found that the overall scale and subscales of C-SPADI had good responsiveness. It was sensitive to functioning status changes of patients who had received shoulder arthroscopic surgery. The values of ES and SRM were slightly larger than those in other study.^[12] This is possibly explained by the fact that the treatment our patients received was arthroscopic surgery, and surgical operation as well as conservative treatment was included in other studies, resulting in the different improvement in functional status.

The process of translation and cross-cultural adaptation was successful. There was only 1 tiny modification was made in item 7. The weight was accustomed to measure in unit called jin in China while it was always measured in unit called pound in Western countries. Hence, "pound" was substituted by "jin" and the corresponding conversion was made.

Strengths of this study are comprehensive assessments of various measurement indexes of the C-SPADI and the longitudinal design of our study. Furthermore, compared with other studies related to other versions of SPADI, the participants were from 2 medical center which could reduce bias during test. However, this study has 3 major limitations. First, the sample was limited in size and not calculated accurately. Second, our translation version of C-SPADI is in simplified Chinese. While mandarin is the official language in China, the many minority groups living here have their own languages. The diversity of ethnicity and cultures has to be taken into consideration in further use. Third, no effect was assessed in the C-SPADI for the patients with shoulder disorders who had received the conservative treatment. This should be dealt with in further research.

5. Conclusion

In summary, the SPADI was translated to Chinese successfully and was confirmed to have good reliability, validity, and

responsiveness. It may be used to assess the functioning status of Chinese patients with symptomatic shoulder disorders in clinical studies or treatments, helping doctors, or researchers to collect required data.

Acknowledgments

We appreciate for the cooperation all the volunteered patients and the staff working in the outpatient center and in-patient department had given us. We also show a sincere gratitude to the 3 experts Prof Zi-ming Wang, Zhi-wei Wang, and An-ren Zhang who helped to assess the C-SPADI, and statistician Mr. Guang-wen Cao who helped for statistical consultation. Special thanks are dedicated to the translators Miss Rui-li Liu, Mr Olivier Assayas, and Mr Jacques Rivette, without whom our research would not be accomplished.

Author contributions

Conceptualization: Wei Wang, Wei Zheng, Wei-dong Xu.
Data curation: Wei Wang, Zhen-yu Jia, Jiao Liu.
Formal analysis: Wei Wang, Jiao Liu, Jin Cui.
Investigation: Wei Wang, Qing-yun Xie.
Methodology: Wei Wang, Zhen-yu Jia, Jiao Liu, Qing-yun Xie.
Supervision: Wei Zheng, Wei-dong Xu.
Validation: Zhen-yu Jia, Qing-yun Xie, Jin Cui, Wei-dong Xu.
Visualization: Wei Zheng, Wei-dong Xu.
Writing – original draft: Wei Wang, Zhen-yu Jia, Wei-dong Xu.
Writing – review & editing: Zhen-yu Jia, Qing-yun Xie.

References

- Luime JJ, Koes BW, Hendriksen JJ, et al. Prevalence and incidence of shoulder pain in the general population; a systematic review. *Scand J Rheumatol* 2004;33:73–81.
- Shan Z, Deng G, Li J, et al. How schooling and lifestyle factors effect neck and shoulder pain? A cross-sectional survey of adolescents in China. *Spine (Phila Pa 1976)* 2014;39:E276–83.
- 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.
- Roach KE, Budiman-Mak E, Songsiridej N, et al. Development of a shoulder pain and disability index. *Arthritis Care Res* 1991;4:143–9.
- Hudak PL, Amadio PC, Bombardier C. Development of an upper extremity outcome measure: the DASH (disabilities of the arm, shoulder and hand) [corrected]. The Upper Extremity Collaborative Group (UECG). *Am J Ind Med* 1996;29:602–8.
- Dawson J, Rogers K, Fitzpatrick R, et al. The Oxford shoulder score revisited. *Arch Orthop Trauma Surg* 2009;129:119–23.
- 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;26:1101–8.
- Pynsent PB. Choosing an outcome measure. *J Bone Joint Surg Br* 2001;83:792–4.
- Zheng W, Li J, Zhao J, et al. Development of a valid simplified Chinese version of the Oxford Hip Score in patients with hip osteoarthritis. *Clin Orthop Relat Res* 2014;472:1545–51.
- Angst F, Goldhahn J, Pap G, et al. Cross-cultural adaptation, reliability and validity of the German Shoulder Pain and Disability Index (SPADI). *Rheumatology (Oxford)* 2007;46:87–92.
- Jamnik H, Spevak MK. Shoulder Pain Disability Index: validation of Slovene version. *Int J Rehabil Res* 2008;31:337–41.
- Tveita EK, Ekeberg OM, Juell NG, et al. Responsiveness of the shoulder pain and disability index in patients with adhesive capsulitis. *BMC Musculoskelet Disord* 2008;9:161.
- Bicer A, Ankarali H. Shoulder pain and disability index: a validation study in Turkish women. *Singapore Med J* 2010;51:865–70.
- Martins J, Napoles BV, Hoffman CB, et al. The Brazilian version of Shoulder Pain and Disability Index: translation, cultural adaptation and reliability. *Rev Bras Fisioter* 2010;14:527–36.

- [15] Marchese C, Cristalli G, Pichi B, et al. Italian cross-cultural adaptation and validation of three different scales for the evaluation of shoulder pain and dysfunction after neck dissection: University of California - Los Angeles (UCLA) Shoulder Scale, Shoulder Pain and Disability Index (SPADI) and Simple Shoulder Test (SST). *Acta Otorhinolaryngol Ital* 2012;32:12–7.
- [16] Christiansen DH, Andersen JH, Haahr JP. Cross-cultural adaptation and measurement properties of the Danish version of the Shoulder Pain and Disability Index. *Clin Rehabil* 2013;27:355–60.
- [17] Alsanawi HA, Alghadir A, Anwer S, et al. Cross-cultural adaptation and psychometric properties of an Arabic version of the Shoulder Pain and Disability Index. *Int J Rehabil Res* 2015;38:270–5.
- [18] Ebrahimzadeh MH, Birjandinejad A, Golhasani F, et al. Cross-cultural adaptation, validation, and reliability testing of the Shoulder Pain and Disability Index in the Persian population with shoulder problems. *Int J Rehabil Res* 2015;38:84–7.
- [19] Membrilla-Mesa MD, Cuesta-Vargas AI, Pozuelo-Calvo R, et al. Shoulder pain and disability index: cross cultural validation and evaluation of psychometric properties of the Spanish version. *Health Qual Life Outcomes* 2015;13:200.
- [20] Phongamwong C, Choosakde A. Reliability and validity of the Thai version of the Shoulder Pain and Disability Index (Thai SPADI). *Health Qual Life Outcomes* 2015;13:136.
- [21] Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res* 2010;19:539–49.
- [22] Beaton DE, Bombardier C, Guillemin F, et al. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine (Phila Pa 1976)* 2000;25:3186–91.
- [23] Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007;60:34–42.
- [24] Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992;30:473–83.
- [25] Xu X, Wang F, Wang X, et al. Chinese cross-cultural adaptation and validation of the Oxford shoulder score. *Health Qual Life Outcomes* 2015;13:193.
- [26] Li L, Wang HM, Shen Y. Chinese SF-36 Health Survey: translation, cultural adaptation, validation, and normalisation. *J Epidemiol Community Health* 2003;57:259–63.
- [27] Altman DG, Schulz KF, Moher D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med* 2001;134:663–94.
- [28] Olofsen E, Dahan A, Borsboom G, et al. Improvements in the application and reporting of advanced Bland-Altman methods of comparison. *J Clin Monit Comput* 2015;29:127–39.
- [29] Perrin C, Khiami F, Beguin L, et al. Translation and validation of the French version of the Western Ontario Shoulder Instability Index (WOSI): WOSI-Fr. *Orthop Traumatol Surg Res* 2017;103:141–9.
- [30] Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res* 2005;19:231–40.
- [31] 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;10:22.
- [32] Wang W, Liu L, Chang X, et al. Cross-cultural translation of the Lysholm knee score in Chinese and its validation in patients with anterior cruciate ligament injury. *BMC Musculoskelet Disord* 2016;17:436.
- [33] Husted JA, Cook RJ, Farewell VT, et al. Methods for assessing responsiveness: a critical review and recommendations. *J Clin Epidemiol* 2000;53:459–68.
- [34] Cohen J. *Applied Multiple Regression/correlation Analysis for the Behavioral Sciences*. Hillsdale: Lawrence Erlbaum Associates; 1983.
- [35] Constant CR, Murley AH. A clinical method of functional assessment of the shoulder. *Clin Orthop Relat Res* 1987;214:160–4.
- [36] Roy JS, MacDermid JC, Woodhouse LJ. A systematic review of the psychometric properties of the Constant-Murley score. *J Shoulder Elbow Surg* 2010;19:157–64.
- [37] Patrick DL, Deyo RA. Generic and disease-specific measures in assessing health status and quality of life. *Med Care* 1989;27(3 Suppl):S217–32.