

BMJ Open Translation, adaptation and measurement properties of an electronic version of the Danish Western Ontario Shoulder Instability Index (WOSI)

Henrik Eshoj,^{1,2} Klaus Bak,^{3,4} Lars Blønd,^{3,5} Birgit Juul-Kristensen^{1,6}

To cite: Eshoj H, Bak K, Blønd L, *et al*. Translation, adaptation and measurement properties of an electronic version of the Danish Western Ontario Shoulder Instability Index (WOSI). *BMJ Open* 2017;7:e014053. doi:10.1136/bmjopen-2016-014053

► Prepublication history and additional material for this paper are available online. To view these files please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2016-014053>).

Received 26 August 2016
Revised 14 March 2017
Accepted 23 May 2017



CrossMark

¹Research Unit for Musculoskeletal Function and Physiotherapy, Department of Sports Science and Clinical Biomechanics, University of Southern Denmark, Odense, Denmark

²Quality of Life Research Center, Department of Haematology, Odense University Hospital, Odense, Denmark

³Sports Clinic, Aleris-Hamlet Parken, Copenhagen, Denmark

⁴Department of Orthopaedic Surgery, Coast Hospital Skodsborg, Skodsborg, Denmark

⁵Department of Orthopaedic Surgery, The Zealand University Hospital, Køge, Denmark

⁶Department of Health Sciences, Institute of Occupational Therapy, Physiotherapy and Radiography, Bergen University College, Norway, Denmark

Correspondence to

Henrik Eshoj; heshoj@health.sdu.dk

ABSTRACT

Objectives To translate and adapt the Western Ontario Shoulder Instability (WOSI) questionnaire into Danish and, to evaluate measurement properties of an electronic Danish WOSI version.

Methods The Swedish WOSI version was used for translation and adaptation into Danish followed by examination of test-retest reproducibility (14-day interval) besides concurrent and construct validity. Concurrent validity was examined by comparing WOSI in paper version with an electronic version, whereas construct validity was examined by comparing WOSI with Numeric Pain Rating Scale (NPRS) and the Oxford Shoulder Score (OSS). Reproducibility was evaluated with Intraclass correlations (ICC), Standard Error of Measurement (SEM), minimal detectable change (MDC) and limits of agreement (LOA). Validity was evaluated with Pearson's (*r*) and Concordance Correlation Coefficients (CCC).

Results 41 subjects (median age 34, range 18–57) were included in the analysis of reproducibility. An ICC of 0.97 (95% CI 0.95 to 0.99) for the total WOSI score was found. SEM was 100.1, resulting in an MDC of 277.5 and LOAs within the range of -246.4 and 308.6. 25 subjects (median age 34, range 18–72) were included in the analysis of concurrent validity obtaining a CCC of 0.96 (95% CI 0.91 to 0.98). Construct validity was investigated in 62 subjects (median age 31, range 18–72) obtaining correlations of 0.83 (95% CI 0.68 to 0.97) (NPRS) and 0.79 (95% CI 0.62 to 0.94) (OSS).

Conclusions An electronic Danish version of WOSI presented excellent test-retest reproducibility and acceptable measurement errors. Also, concurrent validity between paper and electronic version was highly satisfactory as was the construct validity. Surprisingly, though, the NPRS correlated more with WOSI than OSS.

INTRODUCTION

Musculoskeletal complaints are frequent with the glenohumeral joint accounting for up to 8%–13% of all athletic injuries.¹ More specifically, athletic injuries to the shoulder joint often lead to labral-ligamentous lesions inhibiting shoulder stability (SI).² In patients with shoulder instability the ability to participate in sports-related activities is often inhibited resulting in decreased shoulder-related

Strengths and limitations of the study

- Inclusion of patients with shoulder instability from different geographic regions in Denmark improves generalisability.
- There were no missing data due to the use of electronic data collection.
- Strict statistical methods improve the validity of the current findings.
- A factor analysis is warranted to determine the internal validity.
- Responsiveness of the Danish Western Ontario Shoulder Instability was not investigated.

quality of life.^{3–5} Treatment-wise, arthroscopic Bankart repair of the intra-articular shoulder lesions is a commonly used treatment to enhance the stability of the shoulder joint and has been established as the standard procedure when dealing with surgical interventions for SI over the latest decades.^{6,7} Even though arthroscopic SI procedures are considered to be cheap, successful and with relatively low risk of inverse effects,⁸ it is still essential to measure the patient reported treatment effect.^{9,10} Moreover, this is supported by the fact that healthcare professionals tend to overestimate objective patient functions,¹¹ corresponding to poor correlations between the clinical evaluation and subjective impression of patients own health status.^{12,13}

For SI the most frequently used Patient Reported Outcome Measurement (PROM) is the Western Ontario Shoulder Instability (WOSI) questionnaire developed by Kirkley *et al*¹⁴ and is recommended over other PROMs when evaluating treatment effects in SI patients.⁹ Hence, WOSI has been translated into various languages according to international guidelines, and validated for measurement properties nation-wise.^{15–20} In Denmark, an unpublished version of WOSI has been available since 2005 through the

Danish Society for Shoulder and Elbow Surgery (www.dssak.ortopaedi.dk). However, neither was the translation process described nor the questionnaire validated on measurement properties, as opposed to the Swedish version.¹⁷ Furthermore, in large-scale studies (trials, cohorts, surveys, etc.), data are commonly collected electronically, which is time saving and avoids data entry errors.²¹ However, no studies have investigated the validity and reproducibility of an electronic WOSI version yet. Hence, the aim of this study was to cross-culturally translate and adapt the WOSI for use in Denmark besides determining the measurement properties (validity and reliability) of an electronic WOSI version.

MATERIALS AND METHODS

Study design

Cross-cultural translation and adaptation of the WOSI questionnaire for use in Denmark was performed. Furthermore, a prospective longitudinal design was used to examine the measurement properties. This study was exempted for notification to the Health Research Study Board due to the non-invasive/non-treating study design. Though, all subjects gave verbal consent to participate in collecting data for this study.

Cross-cultural translation and adaptation of WOSI

The WOSI questionnaire is designed to measure health-related quality of life in patients with SI and covers four domains with 21 items in all: 'Physical symptoms' (10 items), 'Sport, recreation and work' (four items), 'Lifestyle' (four items) and 'Emotions' (three items). Each item is scored on a visual analogue scale (VAS) ranging from 0 to 100 with a total score of 2100 (with 0 as the level of no trouble).¹⁴ The Swedish WOSI,¹⁷ as valid and reliable as the original version, was used for translation due to the fact that Sweden is cross-culturally and linguistically closer to Denmark than Canada,¹⁴ where from the original version was developed.

A stepwise translation procedure, following international guidelines,^{22 23} was used, including independent forward and backward translations, besides adaptive testing of the final Danish version. Three individuals (one within, two from outside the healthcare system) with Danish as mother tongue produced the forward translation (Swedish to Danish) followed by a backward translation (Danish to Swedish) by two individuals (one within, one outside the healthcare system) with Swedish as mother tongue. Disagreements (wordings, sentence structures, etc.) were handled by a consensus-based method with discrepancies openly discussed between translators until overall agreement was reached. For approval of the final Danish version, the back-translated Swedish version was screened for equivalence to the original Swedish version and accepted by the corresponding author of the Swedish WOSI.¹⁷ Subsequently, to uncover incomprehensible words or sentences a convenient sample of shoulder patients completed the WOSI questionnaire followed by

semistructured interviews until theoretical saturation was reached.

Merging of two Danish versions

The previous Danish translated version of the WOSI questionnaire, initiated in 2005 by the current two coauthors (KB and LB), constituted a translation of the original Canadian WOSI version.¹⁴ According to KB and LB, it was translated following standardised guidelines and approved by the Western Ontario (WO) Group (personal communication). However, the translation process was never described or scientifically tested for measurement properties. Hence, to avoid circulation of two different versions, these two versions were compared for accordance and collapsed into one version. Briefly, merging was performed by thoroughly comparing the two versions for discrepancies/consistencies individually (blinded to each other) by the current authors (KB, LB, HE and BJK). Subsequently, carefully merging of each item was performed with respect to both the original Canadian¹⁴ and the Swedish version.¹⁷ Finally, the merged Danish version was back-translated into English by a professional translator for final revision by one of the coauthors (S Griffin, personal communication) of the original WOSI version.¹⁴

Procedures for examining measurement properties

Study subjects completed a traditional pen and paper WOSI version at the shoulder outpatient clinics or at home for those recruited via phone/email through regular mails receiving a letter with the WOSI paper version and a prepaid return envelope. After completing the paper version, study subjects received an email between 3 and 14 days later, including a hyperlink to an electronic WOSI version and other questionnaires for determining reproducibility besides concurrent and construct validity (figure 1). Further, to improve compliance, subjects received a text message on their phone to remind them to complete the questionnaires and again 2-days later if they had not responded.

Study subjects

Subjects were conveniently recruited as follows: (1) Patients attending shoulder outpatient clinics for a routine check of their shoulder in the subacute phase (approximately 3 weeks postinjury event) following an anterior shoulder dislocation at hospitals situated in the South (South-West Jutland Hospital, Esbjerg and University Hospital Odense) and East (Sports Clinic, Aleris-Hamlet Parken, Copenhagen besides Zealand University Hospital, Department of Orthopaedic Surgery, Køge) Denmark Region; (2) Patients from a postsurgical cohort consisting of 65 patients treated with Bankart repair 5–7 years ago were invited to participate, as well as (3) Patients with self-reported feeling of SI recruited from local advertisements. Recruitment period was from March 2015 to March 2016.

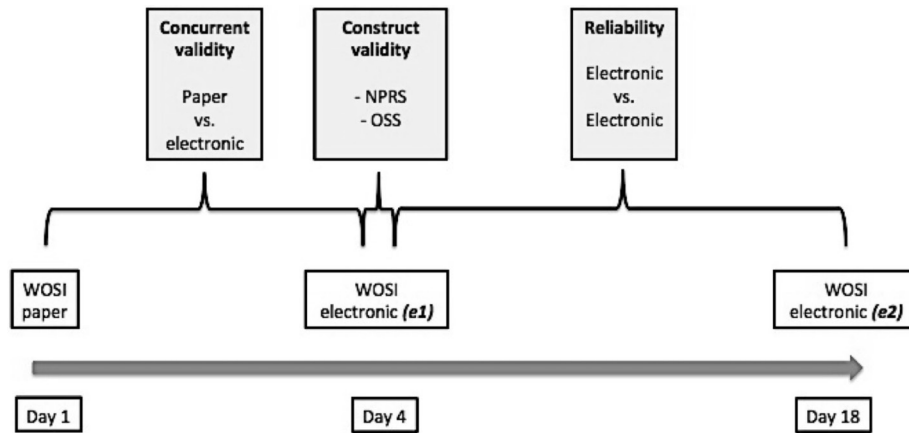


Figure 1 Study design for the examination of test-retest reproducibility, concurrent and construct validity. NPRS, Numeric Pain Rating Scale; OSS, Oxford Shoulder Score.

Eligible subjects were men and women with the following inclusion criteria; (1) Minimum 18 years, (2) Current or previous SI, traumatic dislocations (primary or recurrent), non-traumatic (any subluxation), presurgical or postsurgical conditions (Bankart lesions warranted for or already reconstructed). In case of surgery, subjects were excluded if surgical procedures were performed within the latest 6 months.

Demographic details and descriptive characteristics

Study subjects reported basic demographic details (age, gender, weight and height) and descriptive characteristics (employment status, arm dominance, injury mechanism, previous shoulder dislocations (yes/no), previous surgery (yes/no), as well as use of equipment (computer, tablet, mobile)) when completing the electronic WOSI questionnaire.

Reproducibility of electronic version (test-retest)

Test-retest reproducibility, a measure of consistency and the ability of an outcome to obtain equal results during separate time points,²⁴ was performed between the electronic WOSI version completed at day 4 and day 18 (figure 1). A period of 14 days was considered long enough for subjects not to be able to recall any of their previous responses, but also potentially short enough to remain stable in terms of symptoms, which is essential when evaluating reliability.²⁵ For describing stability of the shoulder condition study subjects completed the following question at day 18 (figure 1): ‘Compared to last time you completed the questionnaire, how would you consider your shoulder function today?’ with one of three possible answers as follows: ‘clearly worse’, ‘largely unchanged’ or ‘clearly improved’.²⁶ Only subjects answering ‘largely unchanged’, representing a stable shoulder condition, were included in the analysis of test-retest reproducibility.

Concurrent validity (electronic versus paper version)

Concurrent validity, which is a measure of how well a particular measure correlates with a previously validated measure,²⁵ was investigated by comparing WOSI

in traditional pen and paper version with an electronic version completed at day 1 and day 4 (figure 1).

For the electronic version a VAS score with an electronic moveable ‘slider’ acting through finger touch (tablet/mobile phones) or at a computer was used with anchor points clearly marked with 0 and 100, representing no pain/problems and extreme pain/problems, respectively. As in the test-retest reproducibility, only study subjects with a stable shoulder condition between day 1 and day 4 were included in the analysis of concurrent validity.

Construct validity (electronic version versus NPRS and OSS)

For construct validity, which is the ability of an outcome measure to accurately evaluate what it is actually intended to evaluate,²⁵ study subjects registered their shoulder pain within the latest 24 hours with the use of a NPRS ranging from 0 to 10 (with ten being worst),²⁷ and completed the Danish validated OSS score.²⁸ OSS is developed to measure outcomes following treatment of shoulder disorders and contains 12 items, individually scored with the use of five response categories, resulting in total OSS score ranging from 12 to 60 (with 12 as the level of no pain or functional limitation). The electronic WOSI version completed at 4 day was compared with the NPRS and OSS, also completed at day 4 (figure 1).

Statistics

For test-retest reproducibility, relative and absolute reliability were calculated. For relative reliability, Intraclass Correlation Coefficients (ICC) and 95% CI was calculated, using a two-way random effects model (2, 1) with single measures and absolute agreement definition.²⁹ ICC was interpreted as follows: <0.40=poor, 0.40 to 0.70=fair to good and >0.75=excellent reliability.³⁰ Measurement errors were interpreted with the use of Standard Error of measurements (SEM), calculated as $SEM = SD \text{ of the mean difference} / \sqrt{2}$,³¹ and minimal detectable change (MDC), calculated by multiplying SD_{diff} with 1.96. Furthermore, SEM and MDC are presented as absolute

values and percentages of the maximal obtainable scores (actual score/2100×100).

For absolute reliability, Bland-Altman plot was used, including 95% upper and lower limits of agreements (LOAs).³² LOAs were estimated by mean difference ± 1.96 times the SD of the difference representing the actual overall difference between measurements one and two. A horizontal line (intersecting zero at the y-axis (black line)) indicates perfect agreement, whereas the dotted horizontal line represents the observed mean difference. The closer the dotted line is to the black line the less disagreement between measurements one and two. This distance was tested for systematic bias by a paired t-test. Furthermore, visual inspection of the Bland-Altman plot was performed to assess funnel effects and systematic bias.

For concurrent validity the Concordance Correlation Coefficient (CCC) was calculated, which is a measure on how well a new set of observations reproduce an original set.³³ Hence, the CCC was used to evaluate how closely the two versions were related to each other and interpreted similar to the ICC as previously described. Furthermore, LOA, SEM and MDC were also calculated.

For construct validity, a Pearson's Product Moments Correlation Coefficient (r) was calculated. A priori the hypothesis was that WOSI would correlate better with OSS than NPRS, since pain is often not the main

symptom in this patient group and that OSS is a PROM specifically developed to measure treatment changes in patients with shoulder disorders.³⁴ However, it was also expected that both the OSS and NPRS reached satisfactory correlations with the Danish WOSI defined as an r value >0.7 .³⁵ The appropriate sample size followed previous recommendations for assessing reliability in health questionnaires.³⁵ Hence, a convenient sample of 50 subjects was determined to be the minimum required number of study subjects. Statistical analyses were performed using SPSS V.23 software, and significance level was set at $p < 0.05$.

RESULTS

In total, 103 subjects were invited to participate (please see flowchart with distribution of subjects fulfilling the requirements for test-retest reproducibility (group I), concurrent (group II) and construct validity (group III), respectively, (figure 2).

The distribution of women/men and median age (range/interquartile range) for group I, II and III was 12/29 and 37 (18–57/19) years, 9/16 and 34 (18–72/20) years besides 22/40 and 31 (18–72/18) years, respectively (please see table 1 for further descriptive characteristics).

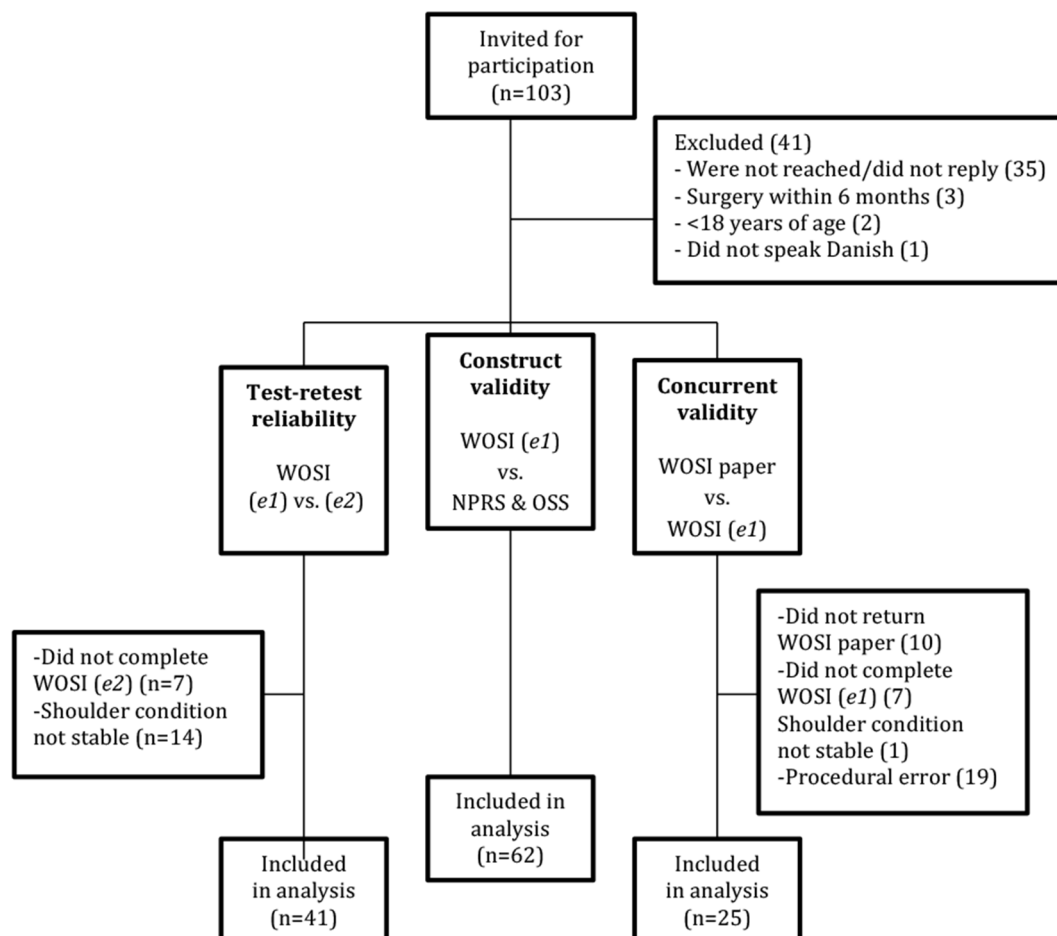


Figure 2 Flow chart for recruitment of subjects and completion of questionnaires.

Table 1 Subject characteristics of participants in group I (test-retest reproducibility), group II (concurrent validity) and group III (construct validity) used for examination of the measurement properties of the Danish WOSI questionnaire

Variables	Group I (n=41)	Group II (n=25)	Group III (n=62)
Sex (women/men)	12/29	9/16	22/40
Age (years), median (range)	37 (18–57)	34 (18–72)	31 (18–72)
Weight (kg), mean (SD)	87.1 (18.1)	86.2 (16.2)	82.5 (17.7)
Height (cm), mean (SD)	181.1 (8.8)	181.0 (8.0)	178.8 (10.0)
Employment, n (%)			
Full-time	24 (59)	18 (72)	35 (55)
Part-time	2 (5)	–	2 (3)
Student	10 (24)	5 (20)	14 (24)
Unemployed/retired	3 (7)	1 (4)	6 (10)
Sick leave	2 (5)	1 (4)	5 (8)
Dominant arm (left/right)	4/38	3/22	34/28
Injured side (left/right)	20/21	12/13	4/58
Injury mechanism, n (%)			
Fall on the arm	22 (53)	14 (56)	30 (48)
Pulling of the arm	8 (19)	4 (16)	14 (22)
Direct external force	1 (2)	–	2 (3)
Other	10 (26)	7 (28)	16 (27)
Previous shoulder dislocation, yes (%)	33 (79)	21 (84)	53 (84)
Previous shoulder surgery, yes (%)	28 (68)	18 (72)	41 (65)
NPRS, latest 24 hours (0–10), mean (SD)	2.41 (2.8)	1.64 (2.1)	2.58 (2.8)
OSS (12–60), mean (SD)	21.1 (9.4)	19.0 (7.8)	21.9 (10.1)
WOSI total score e (1) (0–2100), mean (SD)	735.0 (596.5)	481.8 (414.4)	766.3 (568.4)
Completion of electronic WOSI, n (%)			
Computer	28 (68)	17 (68)	37 (59)
Tablet	3 (7)	2 (8)	6 (10)
Mobile	10 (25)	6 (24)	19 (31)

CM, centimetre; e (1), first electronic completion; OSS, Oxford Shoulder Score; WOSI, Western Ontario Shoulder Instability.

Translation and adaptation

No difficulties regarding translation or merging of the two existing Danish versions were experienced. Moreover, the back-translated versions, whether into Swedish or Canadian (English language), corresponded well with the respective original versions meaning that merging of the two versions was performed with ease. Theoretical saturation, regarding linguistic content and understanding of the Danish WOSI, were reached after interviewing eight shoulder patients. These interviews revealed that only item number two (*‘How much aching or throbbing do you experience in your shoulder?’*) needed further elaboration due to conceptual confusion with the translation of ‘aching’ and ‘throbbing’. This was solved by adding a few clarifying words to the item from the supplemental material (*Explanation of content in questions in WOSI*), which can be found at the end of the questionnaire. Another round of interviews revealed that no further adjustments were needed, and a final approval of a Danish version of the WOSI (Included as supplemental material online) was obtained in January 2015.

Test-retest reproducibility (group I)

Test-retest reproducibility was measured in 41 subjects due to exclusion of 14 (changed shoulder condition from day 4 to day 18), and seven (only one questionnaire completed) subjects (figure 1). The relative test-retest reproducibility was excellent with ICC of 0.97 (95% CI 0.95 to 0.99) for the total WOSI score, and the individual domains varying between 0.93 and 0.96 (95% CI 0.87 to 0.98) (table 2). For the total WOSI score SEM was 100.1, resulting in MDC of 277.5, corresponding to 5% and 13%, respectively, of the maximal obtainable WOSI score. For each of the four domains SEM ranged between 25.7–58.1, and MDC from 71.2 to 161.0, corresponding to 16%–26% of the maximal obtainable score of the individual WOSI domains.

The absolute reliability for the total WOSI score showed no funnel effects or signs of bias in Bland Altman plots, with the 95% LOA within the range of –246.4 and 308.6. Data were approximately equally distributed and closely located to the line of perfect agreement with no statistical test-retest difference ($p=0.15$) (figure 3).

Table 2 Test-retest reproducibility of the Danish WOSI questionnaire (electronic versions) completed within a 14-day interval (n=41)

Domains	WOSI e (1)		WOSI e (2)		Diff.	95% LOA	SEM (%)	MDC (%)	ICC (95% CI)
	Mean±SD	Range	Mean±SD	Range					
Physical symptoms (0–1000)	307.0 (256.1)	0–894	293.2 (268.0)	2–964	13.8 (82.1)	–147.1; 174.7	58.1 (6)	161.0 (16)	0.95 (0.91 to 0.97)
Sports, recreation, work (0–400)	147.5 (131.9)	0–388	135.7 (133.3)	0–396	12.2 (50.1)	–86.0; 110.4	35.4 (9)	98.2 (25)	0.93 (0.87 to 0.96)
Lifestyle (0–400)	136.4 (129.6)	0–381	138.2 (131.1)	0–395	1.8 (36.3)	–69.4; 73.5	25.7 (6)	71.2 (18)	0.96 (0.93 to 0.98)
Emotions (0–300)	144.1 (111.6)	0–300	137.2 (113.1)	0–300	6.8 (39.6)	–70.8; 79.1	28.0 (9)	77.6 (26)	0.94 (0.89 to 0.97)
WOSI total score (0–2100)	735.0 (596.5)	5–1941	704.0 (618.7)	2–2055	31.1 (141.6)	–246.4; 308.6	100.1 (5)	277.5 (13)	0.97 (0.95 to 0.99)

Diff., difference; e (1), first electronic completion; e (2), second electronic completion; ICC, Intra-class Correlation Coefficient; LOA, limits of agreement; MDC, minimal detectable change; WOSI, Western Ontario Shoulder Instability.

Concurrent validity (group II)

Twenty five subjects were included in the analysis of concurrent validity (figure 3) due to exclusion of one (markedly changed shoulder condition from day 1 to day 4), 17 (only one completion, paper or electronic form), and 19 subjects (procedural errors (incorrectly printed version) with VAS scale scores less than 100 mm, or use of the old Danish 2005 version of WOSI), respectively. The concurrent validity showed excellent CCC of 0.96 (95% CI 0.91 to 0.98) (the total WOSI score) and between 0.89–0.96 (95% CI 0.78 to 0.98) (the individual domains) (table 3). For the total WOSI score SEM and MDC was 94.8 and 262.8, respectively. SEM and MDC ranged between 20–56.4 and 55.5–156.4, respectively, for each of the four domains.

Construct validity (group III)

Construct validity was examined on 62 subjects (figure 3). For the total WOSI score the construct validity resulted in correlations of 0.83 (95% CI 0.68 to 0.97) (NPRS) and 0.79 (95% CI 0.62 to 0.94) (OSS), respectively (table 1, group III), while for the individual WOSI domains and NPRS significant correlations varied from 0.65 to 0.85 (95% CI 0.45 to 0.98) (table 4).

DISCUSSION

The WOSI questionnaire was successfully translated and cross-culturally adapted for use in a Danish population of SI subjects. Furthermore, the electronic Danish WOSI had excellent reproducibility including acceptable measurement errors and satisfactory validity. WOSI was translated into Danish, according to international guidelines, by use of bilingual individuals and with no major difficulties experienced. ICCs for the total WOSI score and individual domains all exceeded 0.90, which is regarded as the minimum threshold for use in clinical settings. Furthermore, comparison between paper and electronic WOSI version indicated almost no difference between traditionally (paper) or electronically completion. Finally, for construct validity, significant correlations indicated that the Danish WOSI version is able to measure what it is supposed to measure.

Reproducibility

Test-retest reproducibility of WOSI has previously been reported for both the original and translated versions, though mostly limited to European countries.^{14–20} For the total WOSI score, the current study found an ICC of 0.97 (95% CI 0.95 to 0.99), which was comparable to results by the original version of WOSI¹⁴ and other European translated versions with ICCs of between 0.84–0.95 (95% CI 0.78 to 0.97) indicating that the questionnaire is highly reliable and useful for comparison on both individual and group levels.³⁵ The lower ICC of only 0.84 obtained in the French version²⁰ may be due to the fact that patients were allowed to receive physical therapy treatment during participation in the reliability study, thereby introducing potential bias due to non-stable shoulder conditions

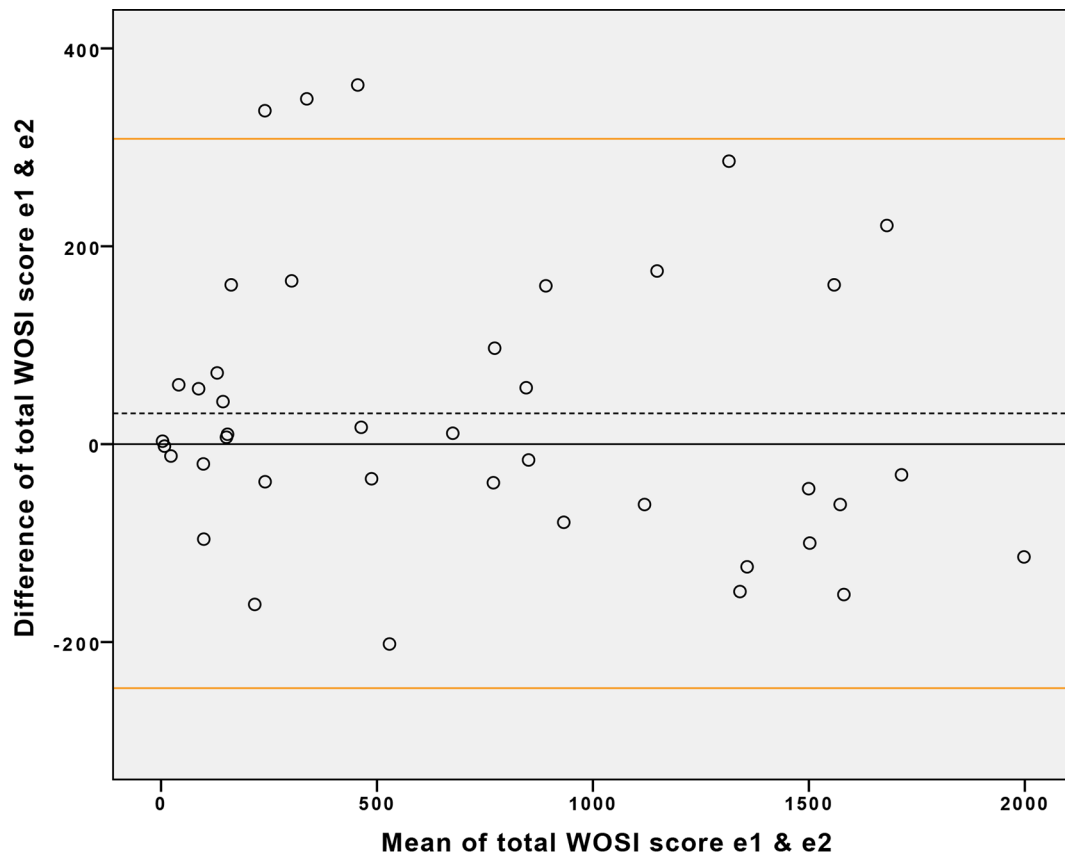


Figure 3 Bland and Altman plot for the reproducibility of the total Western Ontario Shoulder Instability Index (WOSI) score. Dotted line, mean difference between e1 and e2; orange lines, upper and lower limits of agreements (95% LOA). e1, first electronic completion; e2, second electronic completion.

between test and retest. The present study showed SEM of 100.1, which was lower than found by the Dutch and Norwegian versions, but higher than that obtained by the Italian version of WOSI with SEMs of 130.6, 122.4 and 71, respectively.^{15 18 19} The current MDC was 277.5 and thus lower than found by the Dutch¹⁹ and Norwegian¹⁸ versions, but higher than the Italian version¹⁵ corresponding to 362.0, 339.3 and 196.0, respectively. The lower SEM and its resulting MDC in the Italian study could be explained by the use of a test-retest period of only 3 days and use of narrower inclusion criteria (including only patients with primary dislocations).¹⁵ These actions may have generated a more homogenous study group with limited subject variation resulting in smaller SD affecting MDC positively. Differences could also be due to the various ways of calculating SEM³¹ with the Norwegian study¹⁸ calculating SEM as agreement (individual level), whereas the Dutch,¹⁹ Italian¹⁵ and present study calculating SEM consistency (group level). Briefly, SEM agreement takes the systematic differences between test one and two into account (eg, systematic differences between clinicians and their objective measurements), whereas SEM consistency omits these differences.^{35 36} Thus, since investigations of patient reported outcomes are not influenced by objective measurements, the SEM consistency was chosen in the present study.

For agreement evaluation, the present Bland-Altman plot indicated no systematic bias, which is in line with the findings of the Dutch study.¹⁹ The only study to compare LOA with is the Norwegian version¹⁸ with slightly wider LOAs compared with the present version (-339.9; 344.8 vs -246.4; 308.6), which, again, could be due to variance in study populations (eg, occult SI such as superior labral anterior posterior (SLAP) lesions in the Norwegian study versus surgically treated SI in the current study).

Concurrent validity

To our knowledge, this is the first study to report the comparison between a paper and an electronically completed version of WOSI in patients with SI. Previous studies, also investigating the correlation between electronic and paper completed VAS scores, included healthy subjects and showed correlations between 0.86 and 0.99,³⁷ which is similar to the present results. One study, comparing an Ipad (also electronic format) and a paper version in VAS on pain ratings, showed similarly strong correlations.³⁸ Though, the previous studies^{37 38} used an *r-value* to evaluate correlation, whereas the present study uses a CCC. In contrast to using an *r-value*, the CCC does take into account any systematic errors that may exist in the data material obtained.³⁹

Table 3 Concurrent validity between traditional pen and paper versus electronic WOSI version completed within a 3-day interval (n=25)

Domains	WOSI paper		WOSI e (1)		Diff.	95% LOA	SEM (%)	MDC (%)	CCC (95% CI)
	Mean±SD	Range	Mean±SD	Range					
Physical symptoms (0–1000)	186.4±149.8	13–556	192.5±185.0	4–657	6.0±79.8	–150; 162.4	56.4 (6)	156.4 (16)	0.89 (0.78 to 0.94)
Sports, recreation, work (0–400)	83.8±98.4	2–289	96.1±112.5	0–299	12.3±39.7	–65.5; 90.1	28.1 (7)	77.8 (20)	0.92 (0.84 to 0.96)
Lifestyle (0–400)	92.5±101.1	0–319	88.5±96.2	0–330	4.0±28.3	–51.5; 59.5	20.0 (5)	55.5 (14)	0.96 (0.91 to 0.98)
Emotions (0–300)	119.1±98.8	2–289	114.1±111.3	0–300	5.0±38.1	–69.7; 79.7	26.9 (9)	74.7 (25)	0.94 (0.86 to 0.97)
WOSI total score (0–2100)	481.8±414.4	24–1424	491.2±480.7	5–1505	9.3±134.1	–253.5; 272.1	94.8 (5)	262.8 (13)	0.96 (0.91 to 0.98)

CCC, Concordance Correlation Coefficient; Diff., difference; e (1), first electronic completion; ICC, Intra-class Correlation Coefficient; LOA, limits of agreement; MDC, minimal detectable change; WOSI, Western Ontario Shoulder Instability.

Also, in contrast to the current study, the previous study only included healthy, older adults (with low pain levels), thereby limiting the external validity of their findings. Surprisingly, completions of electronic questionnaires have shown not to be influenced by age or previous experience with electronic equipment,²¹ thereby supporting the increasing use of electronic equipment for measuring changes in health status.

Construct validity

The current construct validity of the electronic Danish WOSI was satisfactory when compared with NPRS and OSS. However, the prestated hypothesis of an expected higher correlation between WOSI and OSS, as opposed to NPRS, had to be rejected with NPRS showing slightly higher overall correlation (0.83 vs 0.79) compared with OSS. The reason for this may be that OSS is developed for patients having shoulder surgery other than stabilisation, meaning that some of the items in the OSS³⁴ may not be relevant for subjects with SI as included in the present study.

To our knowledge, construct validity between WOSI and OSS has not previously been tested hampering the possibility for comparison. Indeed, it would have been more appropriate to use PROMs suited specifically for SI (such as the Oxford Shoulder Instability Score (OSIS)). However, the OSIS is not translated nor validated into Danish, and the Danish validated OSS was therefore chosen. Though, WOSI has been tested towards the Disability of the Arm, Shoulder and Hand (DASH),⁴⁰ Constant Murley Score (CMS)⁴¹ and Rowe score.⁴² The present significant correlation was 0.79 between WOSI and OSS, almost identical to a correlation of 0.77 between WOSI and DASH.¹⁶ Usually, lower correlations are found, for example, CMS/Rowe versus patient reported outcomes alone since elements of objective measurements (such as range of motion, strength and/or a specific clinical shoulder examination) are included, which may not reflect the true subjective perception of shoulder function within this patient group.¹⁷

For the construct validity of WOSI towards pain, this study obtained significant correlations of 0.83, almost equal to the Swedish version¹⁷ with correlation of 0.80. Hence, the current correlations between WOSI and pain exceed the correlations between WOSI and the specific shoulder questionnaires, which is surprising since SI symptoms are fluctuant, and often characterised as the feeling of having instability problems rather than pain itself.¹⁷ Nonetheless, pain seems to correlate fairly well with both the individual domains and total score of the WOSI questionnaire in patients with SI, and thereby proves its usefulness. Though, pain and its correlation to WOSI must be carefully interpreted and generalised only to subjects similar to the ones included in the current study.

Table 4 Construct validity between WOSI, NPRS and OSS (n=62)

WOSI (electronic)	NPRS (0–10) <i>r</i> (95% CI)	OSS (12–60) <i>r</i> (95% CI)
Domains		
Physical symptoms (0–1000)	0.85* (0.71 to 0.98)	
Sports, recreation, work (0–400)	0.79* (0.63 to 0.95)	
Lifestyle (0–400)	0.77* (0.61 to 0.94)	
Emotions (0–300)	0.65* (0.45 to 0.84)	
Total score (0–2100)	0.83* (0.68 to 0.97)	0.79* (0.62 to 0.94)

*Significant correlations ($p < 0.01$).

NPRS, Numeric Pain Rating Scale; OSS, Oxford Shoulder Score; *r*, Pearson's Correlation Moments Coefficient; WOSI, Western Ontario Shoulder Instability.

Methodologically considerations

The current study had some limitations and strengths. First of all, the use of the Swedish WOSI for translation, instead of the original Canadian version, may limit the validity of the Danish version due to loss of important cross-cultural and linguistic aspects between Canada and Denmark. However, since the Swedish WOSI was validated and accepted by the WO group before being used for translation in this study and that the Danish WOSI was found comparable to the original Canadian WOSI by the same WO group we believe that the impact of this has been minimal.

Another weakness is the lower sample size of 41 subjects, not meeting the sample size of at least 50 subjects as recommended in reproducibility studies.²⁵ The reason for the lower sample size may be the relatively long test-retest reproducibility period of 14 days increasing the risk of patients experiencing changes in shoulder function and thus not eligible to be included in a test-retest reproducibility analysis. Other reasons for reaching the predefined sample size were procedural errors with patients completing either the unofficial paper version of the Danish WOSI from 2005 or that VAS scales were not 100 mm. However, the present correlations were significant and relatively high with narrow intervals. Therefore we do not see this as a serious weakness in the current study.

The study strengths were the use of standardised guidelines for translating and adapting questionnaires across countries, besides the inclusion of a heterogeneous study group representing hospitals from different geographic populations in Denmark increasing generalisability. Moreover, the use of electronic questionnaires resulted in no missing data nor any technically incorrectly completed items (eg, as is the case with self-administered paper questionnaires). Finally, the use of strict statistical methods increases the validity of the current results.

Perspectives

It can be quite a challenge to keep track of, especially, young and active individuals such as SI patients. However, the use of electronic PROMs provides health-care personnel with the opportunity to collect repeated

measurements with ease. Furthermore, in Scandinavian, validated WOSI questionnaires are now available in Sweden, Norway and Denmark, which allow collaboration projects within Scandinavian SI patients.

CONCLUSION

A Danish version of the WOSI questionnaire has been linguistically and cross-culturally validated for use in a Danish population with SI, and revealed satisfactory reproducibility, concurrent and construct validity. Furthermore, the Danish version is user friendly and can be easily administered electronically thereby meeting today's demands for electronic media usage.

Acknowledgements We would like to thank the participating hospitals and their staff for recruiting study subjects.

Contributors HE and BJK conceived and designed the study. HE, KB and LB recruited the study subjects. HE and BJK planned the statistical analyses. HE performed the statistical analyses. HE, KB, LB and BJK interpreted the results. HE drafted the manuscript with BJK, KB and LB contributing to the manuscript. All authors have read and approved the final manuscript. HE is the guarantor.

Funding Region of Southern Denmark's Research fund, The Danish Rheumatism Association

Competing interests None declared.

Patient consent Obtained.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2017. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

REFERENCES

1. Stevenson JH, Trojjan T. Evaluation of shoulder pain. *J Fam Pract* 2002;51:605–11.
2. Hauser RA. Prolotherapy: A Non-Invasive Approach to Lesions of the Glenoid Labrum; A Non-Controlled Questionnaire Based Study. *Open Rehabil J* 2013;6:69–76.

3. Matsen FA, Zuckerman JD. Anterior glenohumeral instability. *Clin Sports Med* 1983;2:319–38.
4. Robinson CM, Howes J, Murdoch H, *et al.* Functional outcome and risk of recurrent instability after primary traumatic anterior shoulder dislocation in young patient. *J Bone Joint Surg Am* 2006;88:2326–36.
5. van Kampen DA, van den Berg T, van der Woude HJ, *et al.* Diagnostic value of patient characteristics, history, and six clinical tests for traumatic anterior shoulder instability. *J Shoulder Elbow Surg* 2013;22:1310–9.
6. Randelli P, Ragone V, Carminati S, *et al.* Risk factors for recurrence after Bankart repair: a systematic review. *Knee Surg Sports Traumatol Arthrosc* 2012;20:2129–38.
7. Jain NB, Higgins LD, Losina E, *et al.* Epidemiology of musculoskeletal upper extremity ambulatory surgery in the United States. *BMC Musculoskelet Disord* 2014;15:4.
8. Noud PH, Esch J. Complications of arthroscopic shoulder surgery. *Sports Med Arthrosc* 2013;21:89–96.
9. Plancher KD, Lipnick SL. Analysis of evidence-based medicine for shoulder instability. *Arthroscopy* 2009;25:897–908.
10. Simmen BR, Angst F, Schwyzer HK, *et al.* A concept for comprehensively measuring health, function and quality of life following orthopaedic interventions of the upper extremity. *Arch Orthop Trauma Surg* 2009;129:113–8.
11. Lieberman JR, Dorey F, Shekelle P, *et al.* Differences between patients' and physicians' evaluations of outcome after total hip arthroplasty. *J Bone Joint Surg Am* 1996;78:835–8.
12. Koran LM. The reliability of clinical methods, data and judgments (first of two parts). *N Engl J Med* 1975;293:642–6.
13. Koran LM. The reliability of clinical methods, data and judgments (second of two parts). *N Engl J Med* 1975;293:695–701.
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;26:764–72.
15. Cacchio A, Paoloni M, Griffin SH, *et al.* Cross-cultural adaptation and measurement properties of an Italian version of the western Ontario shoulder instability index (WOSI). *J Orthop Sports Phys Ther* 2012;42:559–B6.
16. Hofstaetter JG, Hanslik-Schnabel B, Hofstaetter SG, *et al.* Cross-cultural adaptation and validation of the German version of the western Ontario shoulder instability index. *Arch Orthop Trauma Surg* 2010;130:787–96.
17. Salomonsson B, Ahlström S, Dalén N, *et al.* The Western Ontario shoulder instability index (WOSI): validity, reliability, and responsiveness retested with a Swedish translation. *Acta Orthop* 2009;80:233–8.
18. Skare Ø, Liavaag S, Reikerås O, *et al.* Evaluation of Oxford instability shoulder score, Western Ontario shoulder instability index and euroqol in patients with SLAP (superior Labral Anterior posterior) lesions or Recurrent Anterior dislocations of the shoulder. *BMC Res Notes* 2013;6:273.
19. Wiertsema SH, de Witte PB, Rietberg MB, *et al.* Measurement properties of the Dutch version of the western Ontario shoulder instability index (WOSI). *J Orthop Sci* 2014;19:242–9.
20. Gaudelli C, Balg F, Godbout V, *et al.* Validity, reliability and responsiveness of the French language translation of the western Ontario shoulder instability index (WOSI). *Orthop Traumatol Surg Res* 2014;100:99–103.
21. Greenwood MC, Hakim AJ, Carson E, *et al.* Touch-screen computer systems in the rheumatology clinic offer a reliable and user-friendly means of collecting quality-of-life and outcome data from patients with rheumatoid arthritis. *Rheumatology* 2006;45:66–71.
22. Beaton DE, Bombardier C, Guillemin F, *et al.* Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine* 2000;25:3186–91.
23. 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.
24. Portney LG, Watkins MP. *Foundations of clinical research: applications to practice*. 2nd edn. Prentice Hall: Upper Saddle River, 2000.
25. Mokkink LB, Terwee CB, Patrick DL, *et al.* The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010;63:737–45.
26. Kamper SJ, Ostelo RW, Knol DL, *et al.* Global perceived effect scales provided reliable assessments of health transition in people with musculoskeletal disorders, but ratings are strongly influenced by current status. *J Clin Epidemiol* 2010;63:760–6.
27. Downie WW, Leatham PA, Rhind VM, *et al.* Studies with pain rating scales. *Ann Rheum Dis* 1978;37:378–81.
28. Frich LH, Noergaard PM, Brorson S. Validation of the Danish version of Oxford Shoulder score. *Dan Med Bull* 2011;58:A4335.
29. Rankin G, Stokes M. Reliability of assessment tools in rehabilitation: an illustration of appropriate statistical analyses. *Clin Rehabil* 1998;12:187–99.
30. Fleiss JL. *Reliability of Measurement, in The Design and Analysis of Clinical Experiments*. Hoboken, NJ, USA: John Wiley & Sons, Inc, 1999.
31. de Vet HC, Terwee CB, Knol DL, *et al.* When to use agreement versus reliability measures. *J Clin Epidemiol* 2006;59:1033–9.
32. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1:307–10.
33. Lin LI. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 1989;45:255–68.
34. Dawson J, Fitzpatrick R, Carr A. Questionnaire on the perceptions of patients about shoulder surgery. *J Bone Joint Surg Br* 1996;78:593–600.
35. 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.
36. Streiner DL, Norman GR. *Health measurement scales*. Oxford Medical Publication, 2003.
37. Jamison RN, Gracely RH, Raymond SA, *et al.* Comparative study of electronic vs. paper VAS ratings: a randomized, crossover trial using healthy volunteers. *Pain* 2002;99:341–7.
38. Bird ML, Callisaya ML, Cannell J, *et al.* Accuracy, Validity, and reliability of an electronic visual Analog Scale for pain on a Touch Screen Tablet in healthy older adults: a clinical trial. *Interact J Med Res* 2016;5:e3.
39. Vaz S, Falkmer T, Passmore AE, *et al.* The case for using the repeatability coefficient when calculating test-retest reliability. *PLoS One* 2013;8:e73990.
40. 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.
41. Constant CR, Gerber C, Emery RJ, *et al.* A review of the Constant score: modifications and guidelines for its use. *J Shoulder Elbow Surg* 2008;17:355–61.
42. Rowe CR, Patel D, Southmayd WW. The Bankart procedure: a long-term end-result study. *J Bone Joint Surg Am* 1978;60:1–16.