Cross-cultural adaptation of the Norwegian version of the spinal stenosis measure

Elisabeth Thornes · Margreth Grotle

Received: 16 July 2007 / Revised: 27 November 2007 / Accepted: 16 December 2007 / Published online: 12 January 2008
© Springer-Verlag 2008

Abstract In order to satisfy the need of a tool for assessing the treatment of patients with degenerative lumbar spinal stenosis, an evaluation was made of the reliability, construct validity, and responsiveness of the Norwegian version of Spinal Stenosis Measure (SSM, original by Stucki). This study was a part of a prospective, cohort study. About 75 patients referred for surgery for spinal stenosis participated in the study. A subsample of 30 patients answered the questionnaire twice, test and retest, with at least one week in between. The SSM was translated according to the Guillemin criteria. Reliability was assessed by Bland and Altman's repeatability, intraclass correlation coefficient (ICC) and the coefficient of variance (CV). Internal consistency was assessed by Cronbach's alpha. Construct validity was analysed by correlation analyses. Responsiveness was calculated by the effect size. The reliability between test and retest scores was good for all three subscales of SSM as the ICC-values were above 0.9 and the CVs were below 15%. Cronbach's alpha was above 0.8. The correlation analyses showed high correlation between scales that assessed the same construct, and low to moderate correlation between scales that assessed different constructs. Large effect sizes were found in all the SSM subscales with effect sizes ≥ 1.2 . The Norwegian SSM version has added a highly useful tool for assessing the disease specific status and outcome after treatment in

patients who suffer from degenerative lumbar spinal stenosis.

Keywords Lumbar spinal stenosis · Norwegian version of the spinal stenosis measure · Reliability · Validity · Responsiveness

Introduction

Degenerative lumbar spinal stenosis is the most common lower back disease in the elderly population [1]. Decompression by surgery is often the recommended treatment. Despite many studies, there has been and still is uncertainty regarding the outcome of surgery. Turner et al [16] conducted a meta-analysis of published literature and reported a success rate ranging from 26 to 100%. The reasons for this disparity may include the facts that some of the studies have been retrospective in design, and that different outcome measures have been used. The outcomes have usually been related to the degree of change between pre- and postoperative scores, calculated with different measurement tools, for example the degree of pain measured with a VAS-score or the degree of disability change measured by a disability index.

Gerald Stucki and coworkers [12] have developed a disease-specific self-report measure – the Spinal Stenosis Measure (SSM)—to survey the symptom severity and functional disability related to spinal stenosis as a disease. To evaluate any treatment made to reduce these aspects, the measurement also includes one scale considering the degree of satisfaction after treatment. Stucki describes how the measurement was developed, and the items in each scale were chosen, by using a judgement approach based both on a literature review and on a consensus of an expert panel

E. Thornes (✉)
Department of Orthopedics, Martina Hansens Hospital,
Postbox 23, Donskiv8, Oslo, Gjetttum 1346, Norway
e-mail: elisabeth.thornes@mhh.no

M. Grotle
Diakonhjemmet Sykehus, National Resource Centre
for Rehabilitation in Rheumatology, 0319 Oslo, Norway

consisting of four rheumatologists, an orthopaedic surgeon who specialized in spine surgery, and a behavioural scientist. When the test-retest reliability of the measurement was assessed, the Spearman's rank coefficient ranged from 0.94 (symptom severity and function scale) to 0.96 (satisfaction scale). The validity of the questionnaire was considered to be good as the hypothesized relationship between the questionnaire's three subscales and the external criteria were as expected. They calculated the relative responsiveness to be higher for the symptom severity scale and the physical functional scale than other scales measuring the same aspects of the disease (Sickness Impact Profile and the Roland Scale). Tuli and coworkers [15] have considered the measure validity for overall patient success in lumbar stenosis surgery by using the SSM. Their conclusion is that the threshold values for each of the three scales were similar to previously established values (a change score for symptom severity: 0.46, physical functional: 0.42, and satisfaction: 2.42) and that the most balanced definition of overall success required that a patient achieves an adequate change score in measurement on at least two of the three scales.

During recent years the SSM has been recognized as a useful tool as regards its disease specific content, and it has been used in different studies on degenerative lumbar spinal stenosis [7,10,14]. SSM has been used as the primary outcome measure to evaluate both traditional decompression surgery [9] and the outcome of a new surgical method: "the X STOP Interspinous Process decompression System" [18]. To our knowledge, the SSM exists in English [15] and Slovenian versions [7].

The purpose of this study was to describe the process used to translate and adapt the SSM into Norwegian and to test the Norwegian version of the SSM in terms of test-retest reliability, construct validity and responsiveness among Norwegian patients who had undergone surgery for their spinal stenosis.

Methods

This methodology study was part of a prospective, cohort study carried out on patients with degenerative lumbar spinal stenosis who underwent surgery at a Norwegian hospital. The study was approved by The Committee for Medical Research Ethics in Health Region South in Norway. All patients got written and oral information about the study and gave their written consent to participate.

Questionnaire

The SSM is a self-administered questionnaire developed by G. Stucki and coworkers, published in 1995/1996. The

questionnaire has had several names; Swiss Spinal Stenosis Measure, Zurich Claudication Questionnaire and Brigham Spinal Stenosis Questionnaire.

SSM consists of three different subscales. The first scale considers the severity of symptoms following the disease. The second scale considers physical function characteristic of the disease, and the third scale concerns the patient's satisfaction after treatment. Each scale consists of a number of questions and, if not more than two responses are missing, calculation provides the unweighted mean as an end score. The symptom-scale is scored from 1 to 5, the function and the satisfaction scales from 1 to 4. "1" represents the best possible score on all the three scales, whereas "5" represents the worst possible score on the symptom-scale and "4" the worst possible score on the function- and the satisfaction-scales.

Translation

In order to use the SSM in Norway, we first had to translate and adapt the English version into Norwegian. The translation and adaptation followed international guidelines [2,8]. Two Norwegian people, who were independent and bilingual, translated the English version into Norwegian. They had different backgrounds, one working in offshore and one in healthcare. Two people with English as their native language translated the Norwegian version back to English. Again these two people had different backgrounds, one from healthcare and one a professional translator. Thereafter we compared all versions to discuss and agree on how the final version of the translated SSM should be adjusted to suit the Norwegian population.

The final version was tested in seven people with low back pain, who were asked to read and answer the questionnaire and to comment on the comprehensiveness. None of these seven people reported any difficulty in understanding the Norwegian translated version of SSM. This final Norwegian version was then subjected to further testing with regard to reliability and validity when used in Norwegian patients with lumbar spinal stenosis.

Patients

The study was carried out at Martina Hansen's Hospital, a hospital localised close to Oslo in Norway. The hospital specializes in elective surgery and treatment of orthopaedic conditions and rheumatism.

The patients were referred to the hospital from primary care, and were informed about the study, and asked to

participate after their first visit at the hospital when they were enlisted for surgery. Most of the patients who became enrolled in the study came from the areas around Oslo; Akershus county, Norway.

Procedure and measurements

SSM was part of a comprehensive questionnaire used in the prospective, cohort study. In this questionnaire the patients reported sociodemographic data, medical history, and use of painkillers/sedative medication. They also reported pain intensity in their lower back and/or leg on a 100 mm graded, visual analogue scale (VAS). Functional disability was measured by the Oswestry Disability Index (ODI) [5,6] and psychosocial factors were measured by the Fair Avoidance Beliefs Questionnaire (FABQ) [17], Self-Efficacy Scale (SES) [11], and Hopkins Symptom Check List 25 (HSCL25) [4].

The patients answered the comprehensive questionnaire twice before the surgery (mean time between the two answers was 8.3 weeks) and twice after the surgery (6 weeks and 3 months after surgery).

Reliability

A subsample of 30 patients were asked to participate in the test–retest the procedure, and answered the SSM twice, first before their visit to the hospital 3 months after surgery, and second not later than 1 week after this consultation. At this point after the surgery, the patients' condition was assumed to have become stable. The patients returned the second completion of the SSM by mail.

Validity

To assess the construct of the SSM a discriminate approach was used. Here, we assumed that variables representing the same dimension of the disease would be more correlated than variables representing different dimensions. More specifically, we expected a priori that there would be (1) a strong correlation between the SSM symptom scale and other variables assessing pain, such as the VAS score of leg pain, (2) a strong correlation between the SSM function score and the ODI, and (3) a strong correlation between the SSM satisfaction scale and variables assessing pain (VAS) and disability (ODI) after surgery. Further, we expected low to moderate correlation between the SSM subscales and psychosocial variables such as the FABQ2, SES, and the HSCL25.

Responsiveness

Responsiveness can be assessed in a variety of ways. In this study, we used the same method as in the original study of Stucki et al. [12], which was calculation of the effect size (ES). ES is an account that informs to what extent the measurement is sensitive to clinically important changes. The responsiveness of outcome measures might depend on several factors such as the patient population studied, the type of intervention, the timing of data collection, quantification of the construct of change, and selection of external criteria.

Table 1 Demographic and clinical characteristics of the patients ($n = 75$)

	<i>n</i> (%)
Sex	
Female	42 (56)
Male	33 (44)
Civil status ^a	
Living with someone	56 (76.7)
Living alone	17 (23.3)
Smoking ^a	
Smokers	15 (22.7)
Non-smokers	51 (77.3)
Total years of education ^a	
More than 12-year education	23 (31.1)
Less than 12-year education	51 (68.9)
Working status	
Out of work	52 (69.3)
At work	23 (30.7)
Reason for not working ^a	
Retirement pensioner	39 (68.4)
Backpainproblems	14 (24.4)
Other diseases	4 (7)
Earlier backsurgery ^a	11 (14.7)
Comorbidity	48 (64.9)
Mean SD	
Age (years)	68.4 (10.4)
Years with lumbar pain (years)	14.5 (15)
Backpain (VAS 0–100)	69.5 (23.4)
Legpain (VAS 0–100)	77.6 (17.5)
Oswestry disability Index (0–100)	38.7 (17.5)
Self-efficacy (1–7)	2.8 (1.0)
Hopkin symptom check list 25 (1–4)	1.5 (0.4)
Fear avoidance beliefs questionnaire, physical activity subscale (0–24)	14 (6.4)

Mean values are presented with SD, frequencies in number (n) and percentage (%)

^a Not replied by all patients

Table 2 Analyses of reliability

	SSM symptoms (1–5)	SSM function (1–4)	SSM satisfaction (1–4)
Test, mean (95% CI)	2.42 (2.15–2.7)	1.69 (1.49–1.9)	1.94 (1.66–2.23)
Retest, mean (95% CI)	2.39 (2.12–2.65)	1.74 (1.51–1.97)	1.96 (1.67–2.26)
Difference test/retest, mean (95% CI)	0.04 (–0.07–0.14)	–0.05 (–0.15–0.5)	–0.02 (–0.16–0.12)
Chronbachs alpha	0.96	0.94	0.94
ICC (1.1) (95% CI)	0.92 (0.85–0.96)	0.89 (0.79–0.95)	0.89 (0.76–0.94)
Repeatability (MDC) (in each scales value)	0.55	0.55	0.73
Coefficient of variance (%)	8	12	13

Statistical analysis

The Statistical Package of Social Science (SPSS), versions 12.0 and 14.0, was used to analyse the data. Categorical data are presented as frequencies and percentages. The numerical data are presented as mean and standard deviations (SD). In the test–retest data, mean and 95% confidence intervals (95% CI) were calculated both for the test- and the retest-sample, respectively, and for the difference between the two data-samples. Differences between the data-samples were compared by paired *t*-tests.

To establish the reliability, recommendations by Bland and Altman [3] were followed. In addition to the Intraclass correlation coefficient (ICC 1.1), reliability was assessed both by repeatability or minimal detectable change (MDC) and by the coefficient of variance (CV). To visualize the repeatability, a plot of the difference between first and second responses on SSM against the mean of the sum scores was constructed for each subscale of the questionnaire. The internal consistency of the questionnaire was assessed by Cronbach's alpha. The construct validity was assessed by bivariate correlation analyses, using Spearman's rank coefficient.

Effect Size (ES) was calculated by dividing the mean change by the standard deviation of the baseline scores. A higher ES indicates greater responsiveness and indicates the measurement's ability to catch a change that accounts for the intervention and is not a coincidence. An effect size of "0.8" or more was considered to be large, between "0.4" and "0.8" to be moderate, and below "0.4" was considered small.

In the original evaluation of the SSM, Stucki et al. stated that the SSM satisfaction subscale has a cut-off point at 2.5, suggesting that patients with a satisfaction subscale score of less than 2.5 are assumed to be satisfied with the intervention. We used this cut-off point to calculate the proportion of patients who were satisfied or not, and the mean change scores of the symptom and function subscales for the satisfied/dissatisfied patients, respectively.

Results

The translators involved in the translation of the SSM to Norwegian conditions agreed on the final Norwegian version. None of the seven first persons in a pilot study who had tried out this version reported problems in understanding either the language or the meaning of the answer to each question.

A total of 75 patients with spinal stenosis were included in this study. The demographic and clinical characteristics of the patients are presented in Table 1.

The results of the reliability analyses are shown in Table 2. There were no statistical differences between the test and retest scores. The test–retest reliability analyses proved to be very good, represented by ICC 1.1-values higher than 0.89, repeatability between 0.55 and 0.73, and CV below 15%. The internal consistency measured by Cronbach's alpha was above 0.94 with all three scales. The plots according to Altman are shown in Fig. 1.

The results of the correlation analyses showed that there was a stronger connection between accounts measuring the same quality of the disease than between those measuring a different quality (Table 3). For example, high correlation (>0.70) was found between SSM physical function and ODI scores, and between the SSM satisfaction and the variables assessing pain and disability after surgery. The SSM symptom scale, however, correlated only moderately with leg pain and poorly with back pain. The symptom and physical function subscales correlated poorly with the psychosocial measures, whereas the SSM satisfaction showed moderate correlation with the psychosocial variables such as the FABQ2, SES, and the HSCL25.

There were large effect sizes in all the outcome measures, except the HSCL25 and the FABQ. The symptom severity subscale of the SSM had the highest responsiveness, immediately following leg pain (VAS registered) and the SSM physical function subscale (Table 4).

Most of the patients, 66.2% (*n* = 43), were satisfied with the treatment according to the cut-off point of 2.5, whereas 33.8% (*n* = 22) were dissatisfied. The satisfied patients had

Fig. 1 The plot according to Altman: Individual differences between test and retest are plotted against the mean end score. The *horizontal line* represents mean score of the individual differences, the *dotted lines* represent the 95% limits for repeatability

a mean change of 1.2 (95% CI; 0.9–1.4) and 0.99 (95% CI; 0.69–1.3) on the symptom severity and the physical function subscales, respectively. The dissatisfied patients

had mean change scores of 0.04 (–0.56–0.64) and –0.12 (–0.54–0.31). There were statistically significant differences between the satisfied and dissatisfied patients in the symptom severity scale ($P = 0.01$).

Discussion

The Norwegian version of the SSM appeared to be clearly understood and easily administered by the patients participating in this study. Further, the results showed that the SSM scales had very good test–retest reliability, acceptable construct validity, and good responsiveness.

One of the weaknesses of this study is that we only used medical personnel to translate the measurements. It is now recommended that non-medical people should be asked to translate, to ensure language that is most commonly in use. However, use of medical personnel has made it possible for us to compare the scores on other frequently used measurements, such as the ODI and VAS, with the SSMs three subscales and obtain a wider perspective of both the disease and the influence of surgery. Since there has been little research on patients with degenerative spinal stenosis in Norway, this study has contributed with important knowledge both regarding this new disease-specific SSM tool and other measurements.

The result of the test-retest reliability analyses showed that the SSM had good repeatability when assessing the status of patients with spinal stenosis. These results are comparable to those reported by Stucki et al. [12,13].

By calculating the construct validity, it was confirmed that the SSMs symptom severity subscale and physical function subscale had high correlation with other measurements that measure the same dimensions of the disease and low correlation with those that measure different dimensions. The satisfaction subscale had high correlation with measurements that showed change in pain and functional disability, and moderate with those measuring psychosocial factors. Stucki and coworkers showed by analyses that SSM Satisfaction Scale correlated by 0.66 and 0.68, respectively, to the symptom severity and physical function change in scale points [12]. In our material, the correlations were somewhat higher, 0.78 and 0.59, respectively.

Stucki and coworkers [12] found by calculating relative responsiveness that in patients who were satisfied with surgery, the SSM subscales had a high effect size. This applied both to the physical function subscale (SRM = 1.6), and to the symptom severity subscale (SRM = 1.48). This was confirmed in our material as we found that the disease specific SSM corresponded with the symptom severity subscale, and that the physical function subscale showed higher responsiveness than the other scores used in

Table 3 Construct validity

	SSM symptom severity	SSM physical function	SSM satisfaction
SSM symptom	–	0.25*	0.75**
SSM function	0.25*	–	0.73**
Lower back pain	0.21	0.10	0.63**
Leg pain	0.49**	0.41**	0.75**
Oswestry disability Index	0.31**	0.70**	0.76**
Self-efficacy	–0.24	–0.52**	–0.45**
Hopkin symptom check list 25	0.12	0.17	0.38**
Fear avoidance beliefs questionnaire, physical activity subscale	0.20	0.22	0.51**

The calculations on SSM symptom severity and SSM function are done with baseline responses. The calculations on SSM satisfaction scale is done with the responses 3 months after surgery

* $P \geq 0.05$

** $P \geq 0.01$

Table 4 Responsiveness of outcome measures in the study

	Differences preoperative- 3 months mean SD	ES
SSM symptom (1–5)	0.9 ± 0.8	1.9
SSM function (1–4)	0.7 ± 0.8	1.2
Back pain (VAS 0–100)	25.5 ± 31.8	1.1
Leg pain (VAS 0–100)	29.8 ± 29.2	1.8
Oswestry disability index (1–100)	13.8 ± 16.4	1.1
Self-efficacy (SES) (1–7)	–1.2 ± 1.8	1.0
Hopkin symptom check list 25 (1–4)	0.2 ± 0.3	0.2
Fear avoidance beliefs questionnaire, physical activity subscale (0–24)	4.1 ± 7.4	0.6

ES effect size = mean difference/SD baseline

ES ≥ 0.8 good effect size

ES 0.4–0.8 moderate effect size

this study (apart from leg pain) (VAS-scores). The changes in SSM-scores gave a better picture of the patients' status than for example changes in the ODI or back pain by VAS (Table 4).

Our knowledge of the disease made us choose measurements that assessed different dimensions of the disease such as pain, functional disability, and psychosocial factors. In our material all the measurements included in the comprehensive questionnaire showed statistically significant changes within three months of surgery, although the changes were small in the HSCL25 and the FABQ. However, as an expression of change, each of the pain and disability questionnaires could have been used as an outcome measure.

The advantage of the SSM is that two of the subscales; symptom severity scale and physical function scale combine to provide a picture of the symptoms and disability associated with lumbar spinal stenosis as a disease. The third subscale also detects the patients' own evaluation of the significance of relief from these symptoms and disabilities after treatment. The measure with its three subscales is easy to understand and is answered in ~6 min. With this in mind, and the fact that the SSM also has superior qualities for detecting clinically meaningful changes, the SSM would be an obvious choice when the scientific goal is to handle patients with degenerative lumbar spinal stenosis and assess the effect of treatment in this group of patients.

Conclusion

The cross-cultural adaptation of the Spinal Stenosis Measure to Norwegian has added a highly useful tool for assessing the disease specific status and the outcome after treatment in patients suffering from degenerative lumbar spinal stenosis.

References

1. Arbit E, Pannullo S (2001) Lumbar stenosis: a clinical review. *Clin Orthop Relat Res* 384:137–143
2. Beaton DE, Bombardier C, Guillemin F, Bosi Ferraz M (2000) Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine* 25:3186–3191
3. Bland JM, Altman DG (1999) Measuring agreement in method comparison studies. *Stat Med* 8:135–160
4. Derogatis LR, Lipman RS, Rickels K, Uhlenhuth EH, Cori L (1974) The Hopkins symptom checklist (HSCL): a self-report inventory. *Behav Sci* 1:1–15
5. Fairbank J (1995) Use of Oswestry disability index (ODI). *Spine* 20:1535–1537
6. Fairbank J, Davies JB, Couper J, O'Brien JP (1980) The Oswestry low back pain disability questionnaire. *Physiotherapy* 66:271–273
7. Fokter SK, Yerby SA (2006) Patient-based outcomes for the operative treatment of degenerative lumbar spinal stenosis. *Eur Spine J* 15:1661–1669
8. Guillemin F, Bombardier C, Beaton D (1993) Cross-cultural adaptation of health-related quality-of-life measures: literature review and proposed guidelines. *J Clin Epidemiol* 46:1417–1432
9. Hansraj KK, O'Leary PF, Cammisa FP Jr, Hall JC, Fras CI, Cohen MS, Dorey FJ (2001) Decompression, fusion, and instrumentation surgery for complex lumbar spinal stenosis. *Clin Orthop Relat Res* 18–25
10. Iversen MD, Katz JN (2001) Examination findings and self-reported walking capacity in patients with lumbar spinal stenosis. *Phys Ther* 81:1296–1306
11. Lorig K, Chastain RL, Ung E, Shoor S, Holman HR (1989) Development and evaluation of a scale to measure perceived self-efficacy in people with arthritis. *Arthritis Rheum* 32(1):37–44

12. Stucki G, Daltroy LH, Liang MH, Lipson SJ, Fossel AH, Katz JN (1996) Measurement properties of a self-administered outcome measure in lumbar spinal stenosis. *Spine* 21:796–803
13. Stucki G, Liang MH, Fossel AH, Katz JN (1995) Relative responsiveness of condition-specific and generic health status measures in degenerative lumbar spinal stenosis. *J Clin Epidemiol* 48:1369–1378
14. Tuli SK, Yerby SA, Katz JN (2006) Methodological approaches to developing criteria for improvement in lumbar spinal stenosis surgery. *Spine* 31:1276–1280
15. Tuli SK, Yerby SA, Katz JN (2006) Methodological approaches to developing criteria for improvement in lumbar spinal stenosis surgery. *Spine* 31:1276–1280
16. Turner JA, Ersek M, Herron L, Deyo R (1992) Surgery for lumbar spinal stenosis. Attempted meta-analysis of the literature. *Spine* 17:1–8
17. Waddell G, Newton M, Henderson I, Somerville D, Main CJ (1993) A fear-avoidance beliefs questionnaire (FABQ) and the role of fear-avoidance beliefs in chronic low back pain and disability. *Pain* 52:157–168
18. Zucherman JF, Hsu KY, Hartjen CA, Mehalic TF, Implicito DA, Martin MJ, Johnson DR, Skidmore GA, Vessa PP, Dwyer JW, Puccio ST, Cauthen JC, Ozuna RM (2005) A multicenter, prospective, randomized trial evaluating the X STOP interspinous process decompression system for the treatment of neurogenic intermittent claudication: two-year follow-up results. *Spine* 30:1351–1358