

Test-retest reliability of the Örebro Musculoskeletal Pain Screening Questionnaire and the Situational Pain Scale in patients with chronic low back pain

Emmanuelle Opsommer^a, Roger Hilfiker^b, Barbara Raval-Roland^a, Geert Crombez^c, Gilles Rivier^d

^a HES-SO // University of Applied Sciences Western Switzerland (HESAV) Department of Physical Therapy, Lausanne, Switzerland

^b HES-SO // University of Applied Sciences Western Switzerland (HES-SO Valais-Wallis) Department of Physiotherapy, Leukerbad, Switzerland

^c Department of Experimental Clinical and Health Psychology, Faculty of Psychology and Educational Sciences, Ghent University, Belgium

^d Rehabilitation Centre Clinique Romande de Réadaptation, Service de réadaptation de l'appareil locomoteur, Sion, Switzerland

Summary

OBJECTIVE: To determine the test-retest reliability of the Örebro Musculoskeletal Pain Screening Questionnaire (OMPSQ) and of the Situational Pain Scale (SPS) in patients with chronic low back pain (CLBP).

METHODS: CLBP patients ($n = 30$) who were capable of reading French completed the OMPSQ and the SPS twice with a 1-week interval in one rehabilitation centre in French-speaking Switzerland. To study the test-retest reliability, we calculated intraclass correlation coefficients (ICCs) for the reliability of the overall scores of the two questionnaires.

RESULTS: The ICC for the OMPSQ overall score was 0.89 (95% confidence interval [CI] 0.79–0.95). For the overall scores of the SPS, the ICC was 0.87 (95% CI 0.74–0.93). The standard error of the mean, expressed as percentage of the mean, was 6.6% for the SPS and 10% for the OMPSQ.

CONCLUSIONS: The reproducibility of these two questionnaires in a sample of patients with CLBP is considered good at the overall score level. The French translation of the OMPSQ could be considered as a tool to examine the evolution of psychosocial factors.

Key words: rehabilitation; psychometrics; questionnaires; low back pain

Introduction

Various psychological and social variables play a role in the development, maintenance and exacerbation of back pain problems [1], and to provide suitable and effective treatment for each patient with low back pain remains a daily clinical challenge [2]. Therefore, it is important to take into account individual patient characteristics, including psychological and social factors alongside physical factors [3, 4]. Several tools that explore psychosocial factors [5] are available, amongst them the Örebro Musculoskeletal Pain Screening Questionnaire (OMPSQ), which is a short screening tool [6], and the Situational Pain Scale (SPS) [7], which measures the expected pain in imagined, everyday painful situations. The SPS measures expected pain intensity, and it might potentially be used to predict patients at risk of developing chronic pain problems [5]. The OMPSQ has been validated in several clinical settings in patients with acute and subacute pain consulting a general practitioner or presenting to primary healthcare clinics, and its predictive ability has been documented [8, 9]. The OMPSQ was not designed to measure changes in psychosocial characteristics. However, because psychosocial factors are intervention targets, the assessment of their evolution might provide important insights into mechanisms of patients' outcomes. Furthermore, inclusion of patients' perceptions in outcome measure instruments is recommended [10]. Nevertheless, the usefulness of the OMPSQ as an evaluation tool with chronic pain patients has not yet been explored and in order to do so within a longitudinal cohort study, further validation and assessment of the test-retest reliability in chronic patients is required. The SPS was developed using Rasch methodology [7]; the objective was to develop a unidimensional scale with interval scale properties. There is evidence that this objective was met [7]. However, information about the test-retest reliability with a sufficiently short interval (about 1 week) is lacking. Therefore, the objectives of the study were to examine the test-

Abbreviations

CI	confidence interval
CLBP	chronic low back pain
ICC	intraclass correlation coefficient
IQR	interquartile range
LoA	limits of agreement
MDC	minimum detectable change
OMPSQ	Örebro Musculoskeletal Pain Screening Questionnaire
SEM	standard error of measurement
SPS	Situational Pain Scale

retest reliability of the OMPSQ and SPS in patients with chronic low back pain (CLBP).

Methods

Patients with CLBP who consulted their medical doctor (rehabilitation medicine specialist) between February 2011 and August 2012 in a rehabilitation centre in the French-speaking part of Switzerland were invited to participate. In-

clusion criteria were: (1.) CLBP for more than 3 months verified by a medical doctor; (2.) age between 18 and 65 years, and (3.) ability to read French. Patients were excluded if they had alcohol dependence, a severe psychiatric illness, malignancy, an acute physical problem, infection or a scoliosis with an angle of more than 40°. The study was approved by the local ethics committee and all patients provided informed consent. Questionnaires were administered twice with a 1-week interval.

Table 1: Patient characteristics.	
Characteristic	Number and percentage (if not otherwise stated)
Number of participants	30
Men/women	15/15
Age in years: mean (SD); range	42.3 (11.9); 20–62
Mother tongue	
French	23 (77%)
Other	7 (23%)
Marital status	
Married	20 (67%)
Single	7 (23%)
Separated/divorced	3 (10%)
Life Situation	
Living alone	5 (17%)
Living with partner	24 (83%)
Having children	
Yes	22 (73%)
No	8 (27%)
Highest education level	
Obligatory school	3 (10%)
Vocational education	20 (69%)
High school	2 (7%)
University	4 (14%)
Work situation	
Full-time Job	12 (40%)
Part-time Job	9 (30%)
Part-time job because of pain	1 (3%)
Unemployed because of pain	3 (10%)
Compensation because of pain	4 (13%)
Student	1 (3%)
Number of pain sites	
1	20 (67%)
2	7 (23%)
3	2 (7%)
4	1 (3%)
Pain sites	
Low back pain	30 (100%)
Neck pain	2 (7%)
Upper back pain	6 (20%)
Shoulder pain	3 (10%)
Leg pain	5 (17%)
Pain duration	
12–23 weeks	3 (10%)
24–35 weeks	1 (3%)
36–52 weeks	4 (13%)
More than 52 weeks	22 (73%)
Sick days	
0	20 (67%)
1–7	4 (13%)
>7	6 (20%)

SD = standard deviation

Questionnaires

The OMPSQ [6] has been translated into French, and there is initial validation [11]. A copy of the French version of the OMPSQ is easily available [11]. This instrument was used in this study. The OMPSQ has 25 items, and scores could range in our study from 2 to 210. Higher scores indicated higher risks of poor prognosis. Missing values in this questionnaire were imputed as the mean value of the other items, as recommended by the author of the original questionnaire [6, 12].

The SPS was originally developed in French [7]. A copy of the SPS is available on <http://www.rehab-scales.org/situational-pain-scale.html> (accessed on 2013/09/20). We presented the SPS with the order No. 1. It measures the expected pain in imagined everyday painful situations (e.g., I burn my tongue tasting scorching hot food), and participants have to rate these situations on a verbal pain rating scale. The 18 items of SPS are scored from 0 (not painful), 1 (slightly painful), 2 (moderately painful) to 3 (extremely painful) and a fifth response category corresponding to a “?” response. Situations rated as ‘impossible to estimate’ (“?”) by the participant are encoded as missing data [7]. The score ranges from 0 to 54. Higher scores indicate a worse attitude towards pain. Missing values were treated as 0 [7]. The raw scores were transformed into a linear measure of pain representation with a Rasch analysis and presented as logits (see www.rehab-scales.org). The item locations and the thresholds were anchored to provide the same values as those with the online tool. In this analysis, the missing values were not treated as 0 but taken into

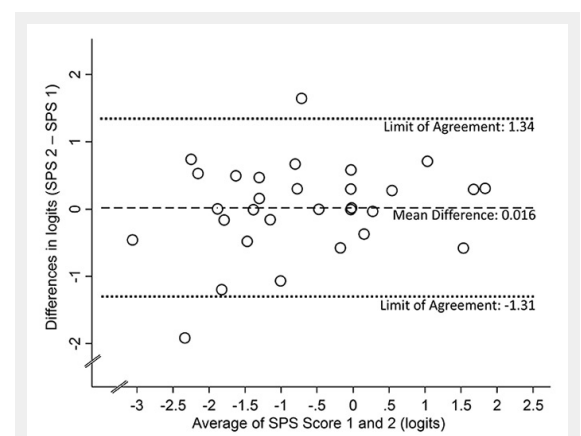


Figure 1

Bland–Altman plot for the Situational Pain Scale (SPS) total score (in logits), with limits of agreements interval (pointed line) of the mean difference (dashed line) between the two assessment occasions.

consideration by the Rasch analysis. The scores in logits (interval scale) of the SPS were used for the analyses.

Patients were required to complete both questionnaires in the presence of a scientific collaborator not involved in the treatment of patients in a quiet room of the rehabilitation centre. Patients were then provided with a second copy of both questionnaires to be completed and posted back 1 week later. We can hypothesise that the bias linked to the variation in the testing situation would decrease the reliability; an important consideration is that this situation “resembles the situation in which the measurement instrument is going to be used” [13].

Sociodemographic data collected were: age, sex, marital status, level of education and professional work status (employed or not), number of missed work days and the clinical data like the duration, location and severity of symptoms, low back pain with or without referred leg pain(s).

Statistical analyses

Data were extracted from the paper questionnaires by user-written software and stored. For further analysis, anonymised data were exported to STATA (StataCorp. LP College Station, TX, USA). The reliability of measures was assessed with the intraclass correlation coefficient ($ICC_{\text{agreement } 2,1}$). The sample size was estimated as follows: with a sample size of 30 and an expected ICC of 0.85, the lower boundary of the 95% confidence interval (CI) would be still above 0.7, which is accepted as a sufficiently high level of reliability [14].

To analyse agreement and reliability on the scale level, we calculated the absolute standard error of measurement (agreement) and the intraclass correlation coefficients ($ICC_{\text{model } 2,1}$ agreement; an $ICC \geq 0.7$ reflecting good reliability [14]). The minimal detectable change (MDC) at the 95% (90%) confidence level was calculated with the standard error of measurement (SEM) as $1.96 \sqrt{2} \text{ SEM}$ ($1.65 \sqrt{2} \text{ SEM}$). The MDC reflects the smallest within-person change in the total score at and above which one can be certain with a given level of confidence that the observed change is above measurement error. There are arguments that the MDC at a 95% confidence level is too stringent, therefore we also report MDC at a 90% confidence level [15]. In addition, we calculated the 95% limits of agreement (LoA) [14] and plotted Bland-Altman plots.

Results

Sixty-eight patients were contacted, of whom 25 were not eligible (CLBP not the main reason for consultation and patients unable to read French were the most frequent reasons

for noneligibility) and 13 did not want to participate or stopped their participation. Of the 30 participants with CLBP included (table 1), 33% had more than one pain site; most had an additional pain site in the upper back. Most patients were living with a partner (83%), and one patient did not respond to this question about the living situation. Sixty-seven percent had had pain for more than 52 weeks. However, 67% were not on sick leave. None received treatment in the hospital between the two measurement timepoints. The median duration between the first and the second questionnaire was 7 days with an interquartile range of 6 to 9 days.

Missing data

Data were generally very complete for the 120 questionnaires. In the OMPSQ, only one item had one missing value with the exception of the two work related questions in the OMPSQ, where eight patients (27%) had missing values. In the SPS, only three items had one missing value; there were 10 items where one patient was not able to respond (response category 5) and for two items there were two patients not able to respond (items 14 and 18).

For both questionnaires, there was no statistically significant difference between the assessments at timepoints 1 and 2 (table 2). For the overall score of the OMPSQ, the ICC was 0.89 (95% CI 0.79–0.95). The ICC for the overall scores of the SPS was 0.87 (95% CI 0.74–0.93). ICC values corresponded to good reliability. Standard error of the measurement ($SEM_{\text{agreement}}$) was 10.12 (10%) for OMPSQ

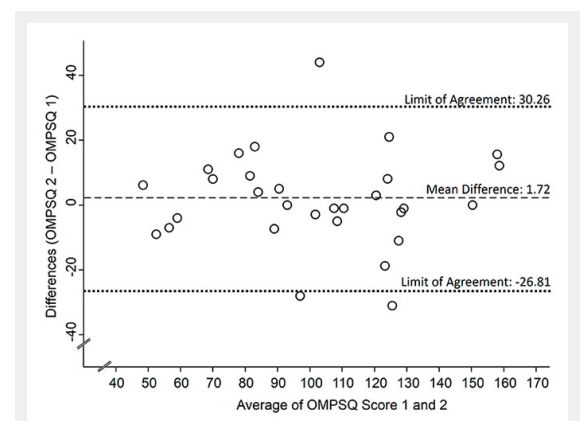


Figure 2

Bland-Altman plot for the Örebro Musculoskeletal Pain Screening Questionnaire (OMPSQ) total score, with limits of agreements interval (pointed line) of the mean difference (dashed line) between the two assessment occasions.

Questionnaire	No. items	Possible range	Mean (SD) first; min to max	Mean (SD) second; min to max	Mean difference (95% CI)	$ICC_{2,1}$ (95% CI)	$SEM_{\text{agreement}}$	SEM (% of mean)	MDC95%	MDC90%
SPS logit	18	0 to 54	-0.693 (1.22); -2.83 to 1.82	-0.677 (1.36); -3.29 to 1.98	0.016 (1.81 to -1.28)	0.87 (0.74 to 0.93)	0.47	6.6%	1.30	1.10
OMPSQ	25	2 to 210	100.85 (31.21); 45 to 153	102.57 (31.18); 48 to 166	1.72 (-3.71 to 7.16)	0.89 (0.79 to 0.95)	10.12	10%	28.1	23.62

CI = confidence interval; ICC = intraclass correlation coefficient; MDC = minimum detectable change; SD = standard deviation; SEM = standard error of measurement

and 0.47 (6.6% of the mean) for SPS. The $MDC_{95\%}$ and $MDC_{90\%}$ values are shown in table 2.

The LoAs from the Bland-Altman plots were -26.81 to 30.26 for the OMPSQ and -1.31 to 1.34 for the SPS. Mean differences were 1.72 for the OMPSQ and 0.016 for the SPS (figs 1 and 2).

Discussion

The results of this test-retest study on patients with CLBP and other musculoskeletal problems indicate that both the French translation of the OMPSQ and the SPS have good test-retest agreement and reliability for use in CLBP. The results apply to the French versions and cannot be generalised to other languages. Even if the population was somewhat different, the estimate of ICC (1,1) obtained from a one-way random effects model was 0.90 (95% CI $0.80-0.95$) in a subsample of 30 patients who completed the Norwegian version of the Acute Low Back Pain Screening Questionnaire (ALBPSQ) twice with an interval of 2 days [16]. Despite the difference between the ALBPSQ and the OMPSQ, with item 5 specifying the painful sites (i.e. 'where do you have pain?') [17] our results are in agreement with this previous study.

The 1-week interval between the two measurement timepoints is a strength of our study because this interval minimises the risk that participants remember how they responded at the first assessment timepoint [18]. Furthermore, the reliability found in our study is a conservative estimate of the true reliability because patients' characteristics might change within 1 week and this introduces a risk of bias towards lower reliability (i.e., an underestimation of the reliability).

A limitation of this study could be its small sample size, but the statistical precision of our study was good enough: the lower end of the confidence interval of the ICC is above the minimum accepted level for reliability of 0.7 [14]. Based on our sample size calculation, there was no need to recruit more patients.

Until now, there were no published data on agreement, reliability and measurement error of both tools with CLBP patients. The reliability of the two questionnaires is comparable to the values found for other assessments in back pain patients [19]. The OMPSQ is designed to identify patients with acute or subacute pain at risk of developing chronic problems. Our study showed that there was no particular problem when patients with chronic pain problems have to fill out the OMPSQ, but further research must evaluate whether the OMPSQ is likewise able to identify patients with chronic pain with a poor prognosis. Furthermore, we present psychometric properties, such as the minimum detectable change, that are important when questionnaires are used as a tool for evaluating interventions. The test-retest reliability is promising, but the usefulness of the OMPSQ as an evaluation tool still awaits further corroboration. For instance, responsiveness is also an important aspect, which we have explored within a longitudinal cohort study (manuscript in preparation). Although the OMPSQ was not designed to be an evaluation tool, its use for evaluation could be interesting as there is evidence that

most psychosocial factors are not stable over time, even in chronic patients [20].

The minimum detectable change should be smaller than the minimum clinically important change [14]. However, for both questionnaires there are no known values for minimum clinically important changes; therefore we cannot determine whether the minimum detectable change is small enough. Furthermore, it must be remembered that the minimal detectable change found in our study might be an overestimate because of the 1-week interval.

In conclusion, given the good reliability observed with the OMPSQ and SPS in this study, both questionnaires can be considered as tools to identify patients at risk of persistent problems, such as the incapacity of return to work, and to examine the evolution of psychosocial factors. However, these issues need further empirical validation. The next step could be to explore the responsiveness and/or the predictive validity of these questionnaires in a longitudinal cohort study of individuals having CLBP.

Acknowledgements: The study was performed at the Rehabilitation Centre, Clinique Romande de Réadaptation, Service de réadaptation de l'appareil locomoteur, Sion, Switzerland. We would like to thank our participants for their time and cooperation. We also thank Miss Virginie Roten who helped in part for carry out the data collection.

Funding / potential competing interests: Details of financial support, national fund, etc.: The study was in part supported by the Swiss National Science Foundation (Grant N° SNF 13DPD6_132178/1) and the HES-SO (RéSaR 07-10_Sagex 23725).

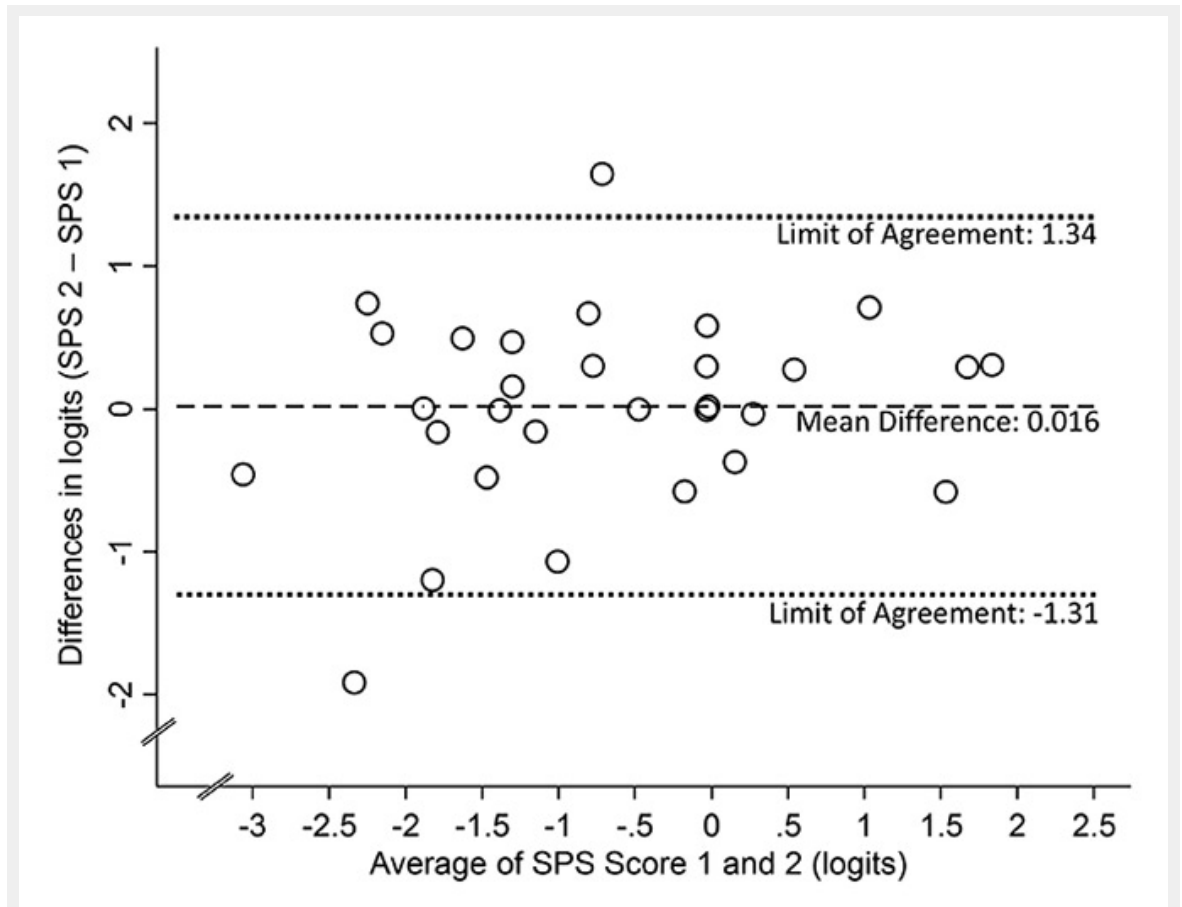
Correspondence: Emmanuelle Opsommer, PhD, HES-SO // University of Applied Sciences Western Switzerland (HESAV), Department of Physical Therapy, Avenue de Beaumont 21, CH-1011 Lausanne, Switzerland, [Emmanuelle.Opsommer\[at\]hesav.ch](mailto:Emmanuelle.Opsommer[at]hesav.ch)

References

- Linton SJ, Shaw WS. Impact of psychological factors in the experience of pain. *Phys Ther.* 2011;91(5):700-11.
- Balague F, Piguet V, Dudler J. Steroids for LBP – from rationale to inconvenient truth. *Swiss Med Wkly.* 2012;142:w13566.
- Hill JC, Fritz JM. Psychosocial influences on low back pain, disability, and response to treatment. *Phys Ther.* 2011;91(5):712-21.
- Steiner AS, Sartori M, Leal S, Kupper D, Gallice JP, Rentsch D, et al. Added value of an intensive multidisciplinary functional rehabilitation programme for chronic low back pain patients. *Swiss Med Wkly.* 2013;143:w13763.
- Turk DC, Melzack R. *Handbook of pain assessment.* New York: Guilford Press; 2010.
- Linton SJ, Boersma K. Early identification of patients at risk of developing a persistent back problem: the predictive validity of the Orebro Musculoskeletal Pain Questionnaire. *Clin J Pain.* 2003;19(2):80-6.
- Decruynaere C. The measure of pain by self-report: use of Rasch analysis. [PhD thesis dissertation]. Belgique; 2007.
- Hockings RL, McAuley JH, Maher CG. A systematic review of the predictive ability of the Orebro Musculoskeletal Pain Questionnaire. *Spine.* 2008;33(15):E494-500.
- Sattelmayer M, Lorenz T, Röder C, Hilfiker R. Predictive value of the Acute Low Back Pain Screening Questionnaire and the Orebro Musculoskeletal Pain Screening Questionnaire for persisting problems. *Eur Spine J.* 2012;21(Suppl 6):S773-84.

- 10 Kempfi C, Laimi K, Salminen JJ, Tuominen R. Perceived relative importance of pain-related functions among patients with low back pain. *J Rehabil Med.* 2012;44(2):158–62.
- 11 Nonclercq O, Berquin A. Predicting chronicity in acute back pain: validation of a French translation of the Orebro Musculoskeletal Pain Screening Questionnaire. *Ann Phys Rehabil Med.* 2012;55(4):263–78.
- 12 Linton SJ. Manual for the Örebro Musculoskeletal Pain Screening Questionnaire: the early identification of patients at risk for chronic pain. Örebro, Sweden: Department of Occupational and Environmental Medicine, Örebro Medical Center; 1999.
- 13 de Vet HCW, Terwee CB, Mokkink LB, Knol DL. Measurement in medicine: a practical guide. Cambridge: Cambridge University Press; 2011.
- 14 Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol.* 2007;60(1):34–42.
- 15 Hopkins WG. Measures of reliability in sports medicine and science. *Sports Med.* 2000;30(1):1–15.
- 16 Grotle M, Vollestad NK, Brox JI. Screening for yellow flags in first-time acute low back pain: reliability and validity of a Norwegian version of the Acute Low Back Pain Screening Questionnaire. *Clin J Pain.* 2006;22(5):458–67.
- 17 Linton SJ, Hallden K. Can we screen for problematic back pain? A screening questionnaire for predicting outcome in acute and subacute back pain. *Clin J Pain.* 1998;14(3):209–15.
- 18 Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures. Statistics and strategies for evaluation. *Control Clin Trials.* 1991;12(4 Suppl):142S–58S.
- 19 Hidalgo B, Gilliaux M, Poncin W, Detrembleur C. Reliability and validity of a kinematic spine model during active trunk movement in healthy subjects and patients with chronic non-specific low back pain. *J Rehabil Med.* 2012;44(9):756–63.
- 20 Burns JW, Kubilus A, Bruehl S, Harden RN, Lofland K. Do changes in cognitive factors influence outcome following multidisciplinary treatment for chronic pain? A cross-lagged panel analysis. *J Consult Clin Psychol.* 2003;71(1):81.

Figures (large format)

**Figure 1**

Bland-Altman plot for the Situational Pain Scale (SPS) total score (in logits), with limits of agreements interval (pointed line) of the mean difference (dashed line) between the two assessment occasions.

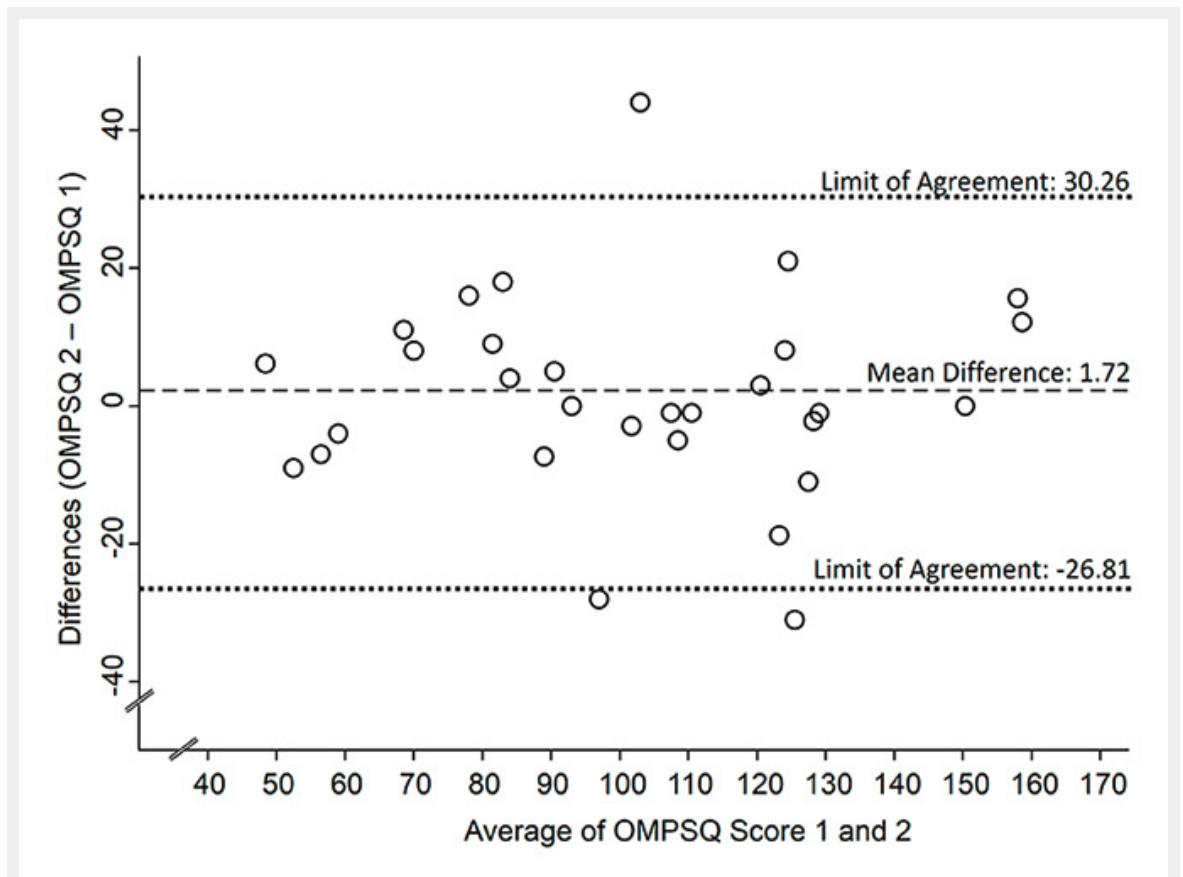


Figure 2

Bland-Altman plot for the Örebro Musculoskeletal Pain Screening Questionnaire (OMPSQ) total score, with limits of agreements interval (pointed line) of the mean difference (dashed line) between the two assessment occasions.