Responsiveness of the Chinese version of the Oswestry disability index in patients with chronic low back pain

Chao Ma · Shaoling Wu · Lingjun Xiao ·
Yunlian Xue

Received: 29 April 2010 / Revised: 2 September 2010 / Accepted: 7 November 2010 / Published online: 26 November 2010
© Springer-Verlag 2010

Abstract The Oswestry Disability Index (ODI) is one of the most widely used questionnaires that assess disability in patients with low back pain (LBP). Responsiveness is both an important psychometric property of an instrument and a key issue for clinicians when choosing suitable outcome measures. The objective of this study was to examine the responsiveness of the Chinese version of the ODI (ODI-Chinese) for subjects with chronic LBP following a physical therapy program. In total, 76 patients with chronic LBP completed the ODI-Chinese, a visual analog scale (VAS) of pain, and the Chinese version of Short Form-36 (SF-36) before and after treatment. All patients also completed a global perception of change Likert scale in condition after the program. The scale was collapsed to produce a dichotomous variable outcome, improved or non-improved. The responsiveness of the instruments was determined using the standardized response means (SRM) and receiver operating characteristics (ROC). After treatment, 56 patients considered themselves to be improved. The SRM of the ODI-Chinese was -1.2 in the improved group and -0.4 in the non-improved group. The area of the ROC curve for the ODI-Chinese was 0.77 (95% CI 0.66 – 0.89). Therefore, the

Chinese version of the ODI is both responsive and appropriate for use in chronic LBP patients after conservative therapy.

Keywords Low back pain · Oswestry Disability Index · Disability · Responsiveness

Low back pain (LBP), which is often with associated leg symptoms, is an important cause of disability and work absenteeism [19]. In China, it has been reported that the 1-year incidence of LBP in workers and teachers was 50%. This increases to 64% in the rural population [14]. Given that the population of China is approximately 1.3 billion, we estimate that more than half a billion Chinese people suffer from LBP. Because LBP is common and has negative effects on a person's daily life, work activities, and quality of life, the level of disability is an important outcome indicator in both clinical and research settings. Outcome measures determine the level of a patient's impairment and disability, choice of therapy, and observation of therapeutic efficacy. To be clinically meaningful, an outcome instrument must be valid, reliable, responsive, and easy to both complete and score [5]. Responsiveness is a critical characteristic of an instrument; it is defined as the ability to detect clinically relevant change over time [1].

Self-reported questionnaires of pain and functional status have become important outcome measures when evaluating patients with LBP. The Oswestry Disability Index (ODI) is one of the most widely-used outcome measures for individuals with LBP. Consequently, it has now been translated into at least 15 different languages [4, 8]. The original version of the ODI was developed in English [7]; the Chinese version was validated recently [16].

C. Ma · S. Wu (✉) · L. Xiao
Pain Treatment Centre of Department of Rehabilitation
Medicine, Sun Yat-sen Memorial Hospital,
Sun Yat-sen University, 107 Yanjiang West Road,
510120 Guangzhou, Guangdong, China
e-mail: wushaolinggz@126.com

Y. Xue
Information Department, Sun Yat-sen Memorial Hospital,
Sun Yat-sen University, Guangzhou, Guangdong, China

Fig. 1 Patient recruitment and follow-up flow diagram

The purpose of this study was to examine the responsiveness of the Chinese version of the ODI in subjects with chronic LBP.

Methods

Participants

One hundred and twenty-nine patients with non-specific, chronic LBP were recruited from our pain-treatment center between January 2009 and October 2009. Subjects were invited to participate if they were at least 18 years of age and had a minimum of 3 months of history with LBP with or without leg pain or neurological signs. Systemic rheumatological disease, gynecological problems, ankylosing spondylitis, tumors, infection, pregnancy, and serious spinal pathology (cauda equina symptoms) were excluded.

Ultimately, 76 patients completed baseline assessments and 6-week follow-up outcomes. These 76 were analyzed in this study; the patient-flow is shown in Fig. 1.

All subjects gave written consent to participate in the study. The Ethics Committee for Medical Research at our institution approved the study.

Procedures

Baseline assessments were performed by research assistants in our outpatient departments. A questionnaire package, consisting of a medical history, the Chinese version of the ODI Version 2.1, a visual analog scale (VAS), and the Chinese version of Short Form-36 (SF-36), was administered to all patients. Following the baseline assessments, subjects were treated with 6 weeks of physical therapy. This included electrotherapy, manual therapy, specific exercises for stabilization, and therapeutic exercises. A small number

of subjects received further intervention by physiotherapists during the study period. At the follow-up time point, both the questionnaires and a 7-point Likert scale were assessed. The assessments were administered by the same assistants that performed them at baseline. The patients completed the questionnaires on their own after the assistants explained the scales.

Instruments

The ODI V. 2.1 is a disease-specific measure developed for the assessment of back pain [8]. It has 10 dimensions: pain intensity, personal care, lifting, walking, sitting, standing, sleeping, sexual activity (if applicable), social life, and traveling. Each dimension has six levels. A score of 0 is used for the least-disabled level; a score of 5 is used for the most-disabled level. The total score is converted to a percentage with a consequent maximum of 100%. Version 2.1 used in this study has been translated and cross-culturally adapted for Chinese patients [16].

A VAS is a 100-mm-long horizontal line that is anchored by word descriptors at each end (no pain and worst pain possible). The patient selects the point on the line that best represents his or her perception of their pain level [10].

The SF-36 is a generic health profile instrument. It includes physical functioning (PF), social functioning (SF), role-physical (RP), role-emotional (RE), mental health (MH), vitality (VT), bodily pain (BP), and general health (GH) [22]. For each of these eight dimensions, item scores are coded, summed, and transformed to a scale of 0 (worst possible health state) to 100 (best). The SF-36 was further combined into two summary scores: physical component score (PCS) and mental health component score (MCS) [24]. The psychometric properties of the Chinese version of the SF-36 are well documented [17].

At 6 weeks of follow-up, patients were asked about their perception of the magnitude of change in both their pain and physical disability since the beginning of their treatment. This retrospective assessment was formulated on a 7-point Likert scale [3]. The scale has four improvement levels (completely better, much better, better, and little better), one no change level (approximately the same), and two worsening levels (a little worse and a lot worse). The levels of global perception of the condition scale were collapsed to produce a dichotomous variable outcome: improved (completely better, much better, and better) and non-improved (a little better, approximately the same, a little worse, and a lot worse).

The standard error of measurement (SEM) is a common method of expressing the imprecision of an instrument. The 95% tolerance interval (95% TI) is the smallest score change that is possible to detect beyond error [13].

Statistical analyses

The responsiveness of the questionnaires was determined by calculating the standardized response means (SRM) and the receiver operating characteristics (ROC). The data were analyzed using the SPSS software (Ver. 11.0 for Windows; SPSS Inc., Chicago, IL, USA). A p value <0.05 was considered to be statistically significant.

The change scores for each instrument were calculated by subtracting the baseline scores from the follow-up scores. Consequently, a negative change indicated an improvement for the ODI and VAS; a positive score indicated an improvement for the SF-36 PCS and SF-36 MCS. All change scores were checked to determine whether they were normally distributed by using the K–S Lilliefors test and by inspection of histograms and box plots of the data. Statistical significance in differences in the change scores in the improved and non-improved subgroups was tested by independent samples t tests. The SRMs were calculated by dividing the mean change by the standard deviation of the mean change scores [9]. Effect sizes of SRMs of 0.2, 0.5, and 0.8 or above are considered to be small, moderate, and large clinical changes, respectively [9].

It has been suggested that the responsiveness of an outcome measure can be considered analogous to the evaluation of a diagnostic test. Using this analogy, the questionnaire is the diagnostic test and the global result represents the gold standard [6]. The ROC curve synthesizes information on both sensitivity and specificity to identify alterations in the condition in agreement with an external, dichotomized result. In this study, the Likert scale was collapsed into an external, dichotomized instrument. Therefore, the ROC curve can be interpreted as the probability of correctly discriminating between the improved and non-improved subgroups. This area theoretically ranges from 0.5 (no accuracy in discriminating improved from non-improved) to 1.0 (perfect accuracy). The point of the ROC curve that is nearest to the upper-left corner of the diagram was used to estimate the minimum change necessary to distinguish between improved and non-improved patients [6]. The cutoff point was found by identifying the point with highest sensitivity and specificity in the ROC analysis that usually takes the point with the highest Youden index.

The standard error of measurement (SEM) was calculated for the whole study population ($n = 76$) with the following formula: $SD \times \sqrt{1 - \alpha}$. Here, the SD is the standard deviation of the pre-treatment score and α is the coefficient of test–retest reliability. The 95% tolerance interval (95% TI) was calculated to be $\pm 1.96 \times \sqrt{2} \times SEM$. The reliability coefficient that we used (0.90) was obtained from previous studies [13].

Table 1 Demographic and clinical characteristics of the subjects ($n = 76$)

Variable	Frequency	Mean $\pm$ SD
Age (years)		42.8 $\pm$ 15.2
Gender		
Male	50 (65.8)	
Female	26 (34.2)	
Occupation		
Laborer	22 (28.9)	
Office worker	30 (39.5)	
Home working	10 (13.2)	
Students	5 (6.6)	
Retired	9 (11.8)	
Pain duration (months)		32.7 $\pm$ 41.6
3 months to 1 year	12 (15.8)	
1–5 years	48 (63.2)	
≥ 5 years	16 (21.0)	

Results

The socio-demographic characteristics for the patients are shown in Table 1. In total, 56 (73.7%) patients were classified as having experienced a health improvement. The number of patients assessed as completely better, much better, or better were 10, 31 and 15, respectively. There were 20 (26.3%) patients who reported their health status as non-improved. The number of patients assessed

as little better, approximately the same, a little worse, and a lot worse were nine, six, four, and one, respectively. Table 2 shows baseline mean scores, follow-up mean scores, SDs, change scores with SD, and SRMs. It also shows the 95% confidence intervals for each instrument by patient subgroup (improved and non-improved). Except for the SF-36 MCS, all of the change scores of the outcome measurements showed significant differences between those reporting improvement and non-improvement. In the improved group, the ODI produced the largest SRM (−1.2), while VAS, SF-36 PCS, and SF-36 MCS generated moderate SRMs (−0.6, 0.7, and 0.5, respectively). In analyses of the non-improved group, the instruments showed small SRMs (−0.4 for ODI, 0.2 for VAS, and 0.2 for SF-36 PCS), except for the SF-36 MCS (−0.1).

Table 3 and Fig. 2 show that the ODI was the most responsive instrument, with an area of 0.77 (95% CI 0.66–0.89). The VAS achieved an area only slightly smaller than the SF-36 PCS at 0.67 and 0.71, respectively. The value for the SF-36 MCS was non-significant. The optimal cutoff points were approximately 5.3 points for the ODI (86.2% sensitivity; 66.3% specificity), 3.0 points for VAS, and 5.0 points for the SF-36 PCS.

The SEM and the 95% TI of the ODI were 4.9 and 13.5, respectively. The 30% change of the baseline in our study was 11.2. We considered this to be the minimal important change (MIC) value for the ODI-Chinese after 6 weeks of physical therapy.

Table 2 Mean baseline and post-treatment change scores with standard deviations (SD) and standardized response mean (SRM) with 95% confidence intervals (CI) in patients reporting improvement or non-improvement

	Mean baseline (SD)	Post-treatment (SD)	Mean changes (SD) (post–pre)	<i>p</i>	SRM	95% CI for the SRM
ODI ^a (0–100)						
Improved	37.5 (15.8)	23.8 (10.2)	−13.7 (11.8)	<0.001	−1.2	−1.6, −0.9
Non-improved	36.8 (14.5)	34.5 (7.2)	−2.3 (6.1)		−0.4	−0.7, 0.0
VAS ^a (0–100)						
Improved	42.8 (21.4)	30.7 (20.5)	−12.1 (19.7)	<0.001	−0.6	−0.9, 0.1
Non-improved	43.6 (23.0)	47.4 (19.2)	3.8 (18.5)		0.2	−0.4, 0.7
SF-36 PCS (0–100)						
Improved	21.8 (9.6)	29.5 (13.1)	7.7 (11.6)	0.048	0.7	0.3, 1.2
Non-improved	22.4 (8.7)	24.0 (7.0)	1.6 (7.5)		0.2	−0.3, 0.6
SF-36 MCS (0–100)						
Improved	46.7 (12.9)	52.1 (9.2)	5.4 (11.6)	0.062	0.5	0.0, 1.0
Non-improved	45.0 (11.7)	43.5 (15.1)	−1.5 (10.3)		−0.1	−0.6, 0.4

ODI Oswestry Disability Index, VAS visual analog scale, SF-36 PCS physical component score of the SF-36, SF-36 MCS mental health component score of the SF-36

The *p* values refer to statistical differences in the change scores between the improved and non-improved groups (independent *t* test). The 95% CIs for the SRMs were calculated by the bootstrapping technique

^a Higher scores on the ODI and VAS indicate lower functional ability

Table 3 Area under the ROC curve (improved vs. non-improved) for each instrument

Instrument	<i>n</i>	Area	95% CI	Optimal cutoff value for change on instrument	Sensitivity at optimal cutoff (%)	Specificity at optimal cutoff (%)
ODI	76	0.77	0.66–0.89	5.3	86.2	66.3
VAS	76	0.67	0.54–0.80	3.0	74.0	58.4
SF-36 PCS	76	0.71	0.58–0.84	5.0	82.6	61.2
SF-36 MCS	76	0.60 (ns)	0.48–0.70	3.6	69.5	51.0

All ROC analyses were statistically significant ($p < 0.05$) except for the one marked ns (not significant)

ODI Oswestry Disability Index, VAS visual analog scale, SF-36 PCS physical component score of the SF-36, SF-36 MCS mental health component score of the SF-36

Discussion

In this study of the patients with LBP, the responsiveness in three frequently used outcome measurements was explored. Applying a global change index as an external criterion, the ODI-Chinese showed large SRMs (−1.2) in the improved group and moderate SRMs (−0.4) in the non-improved group. The area of the ROC curve was 0.77 (95% CI 0.66–0.89). The VAS and the SF-36 PCS showed less favorable results for both of the groups. We considered the MIC value of ODI-Chinese to be the 11.2 point after 6-weeks of treatment with physical therapy in our study.

Instruments for low back pain

There are several generic and disease-specific measures available for evaluating patients with LBP; selecting the proper tool is challenging for most clinicians and researchers. The SF-36 is one of the most widely used generic health questionnaires and is increasingly being used for measuring outcomes in patients with LBP [5, 24]. The two most-common, disease-specific and well-studied instruments for patients with LBP are the ODI and Roland Morris Disability Questionnaire (RMDQ). Studies have found that the ODI was more responsive than RMDQ [9]. The ODI has been recommended by the National Spine Network (NSN) [24]. Furthermore, the ODI is well-suited for use in busy clinical settings that focus on assessing both the effect of chronic LBP and gains after interventions [23].

The SRM of the ODI-Chinese

The responsiveness of the ODI has been compared between a number of different populations and context [3, 5, 10, 12, 15, 24]. These studies have estimated an SRM between 0.3 and 1.2 for the ODI [5, 9, 10, 15, 24]. Our study demonstrated similar responsiveness for the Chinese version of the ODI. The SRM of the ODI-Chinese was −1.2 for the improved group and −0.4 for the non-improved group. These results

Fig. 2 ROC curves of the change scores for ODI, VAS, SF-36 PCS, and SF-36 MCS, using the outcome of global perception of the condition scale (dichotomized as the external criterion). ODI Oswestry Disability Index, VAS visual analog scale, PCS physical component score of the SF-36, MCS mental health component score of the SF-36

show that the SRM of the instrument in the improved group was significantly larger than in the non-improved group, as has been previously demonstrated [3, 10]. In the improved group, the change scores were larger (i.e., produced a larger SRM) than those in the non-improved group. Previous research has found that the patient group, treatment, and timing of data collection can affect the magnitude of change measured by a given instrument [3, 13].

The SRM of the ODI-Chinese was −0.4 for the non-improved group indicated that a number of patients whose health were not improved according to the 7-point Likert scale still had an important reduction in their ODI scores. This may be partly because the patients in the non-improved group actually had some improvement of function after 6 weeks of physical therapy. On the other hand, the

non-improved group according to the 7-point Likert scale includes the conditions of a little better, approximately the same, a little worse, and a lot worse. Also, the conditions of little better and approximately the same were majority in our study. The results of the SRM of the ODI-Chinese for the non-improved group (-0.4) were only slightly less than the SRM of the VAS in the improved group (-0.6). This indicates that the ODI instrument is also sensitive in the non-improved group.

In all scales, there were some inconsistencies between patients' 7-point Likert scale rated self-assessment and the change scores. Guyatt et al. [11] has suggested that patients should be given access to their pretreatment questionnaires before completing the follow-up questionnaire to potentially reduce the number of inconsistencies between self-assessment of change after treatment and scores from the outcome measures.

Comparison of the ODI and the SF-36 for patients with LBP

The issue of generic and disease-specific measures for patients with LBP have been studied previously by many authors [10, 22, 24]. The pain and function questions imbedded in the SF-36 are well suited for assessing patients with LBP. In particular, the wording of the SF-36's physical questions is biased toward lower extremity function. These studies found that the BP, PCS, SF, and PF scales of the SF-36 showed good responsiveness for patients with LBP, whereas the others are less consistent [10, 22, 24]. It is commonly believed that the usefulness of general measures lies in their ability to allow comparisons both between patients with the same condition and between patients with different conditions. Moreover, general measures may be able to identify unexpected side effects from a new treatment. As a result, Taylor et al. [22] suggested that the SF-36 questionnaire can be a useful adjunct in the assessment of patients with LBP.

In our study, the SRM values of the ODI in the improved and non-improved groups were -1.2 and -0.4 , respectively. This was higher than those for the SF-36 PCS, which were 0.7 and 0.2 , respectively. Importantly, only the ODI was able to distinguish significantly between the improved and non-improved groups when considering the confidence intervals around the SRMs. The areas under the ROC curves for the ODI (0.77) were larger than those for the SF-36 PCS (0.71), with a non-significant value for the SF-36 MCS. These results showed that the ODI was more responsive than the SF-36 in the evaluation of patients with LBP. Although there are problems with responsiveness, the SF-36 has some items that are more concrete than the ODI, such as the item of go upstairs. As a result, we can gain more information from the SF-36.

Minimal important change

The cutoff point of ROC analyses has frequently been used to estimate the minimum clinically important difference (MCID) [25]. These reported MCID values for the ODI range from 4.5 to 23 points [2, 3, 18, 24]. The wide variation is a result of the heterogeneity of the studies, the disparate results, and the lack of any clear rationale for integrating them [21]. With a method that takes the point from the highest Youden index of the ROC, we found that the optimal cutoff value of the ODI was 5.3 points. This value was well below the accepted minimum clinically important differences. The use of a retrospective global assessment as an external criterion has been challenged by Norman et al. [20] due to the problem of recall bias. Still, some studies considered that the external criteria used in the analysis permitted the identification of patients whose health had changed over time. However, whether such changes were important either the clinician or the patient could not be ascertained [21]. Considering the apparent confusion in the literature about the distinction between "estimated" and "important" change, it is noteworthy that such change scores should not be interpreted as thresholds for clinically meaningful changes. More importantly, some studies have revealed that the anchor-based approaches do not take measurement precision into account. They also do not necessarily imply statistical significance, which consequently, must be calculated separately [21]. The researchers suggested that the change scores of an instrument should be compared with the SEM and the 95% TI [13]. In our study the SEM and the 95% TI of the ODI were 4.9 and 13.5 , respectively. The study by Hägg et al. [13] had similar results regarding the smallest acceptable score change of the ODI. It suggested that the score change may require an increase exceeding the 95% tolerance interval when used for clinical decision making and power calculation. These methodological problems have led to confusions about how to both interpret change score and decide what change is clinically important. Naturally, statistical significance does not necessarily imply that the change is clinically important. Therefore, an expert panel and a workshop during the "VIII International Forum on Primary Care Research on Low Back Pain" (Amsterdam, June 2006) worked towards creating international agreement regarding minimal important change. Eventually, they reached a reasonable consensus that the proposed a minimal important change (MIC) value of 10 for the ODI. When the baseline score is taken into account, a 30% change was judged to be a clinically meaningful improvement when comparing outcomes before and after treatment for individual patients [21]. Although not a final answer, the proposed MIC value may offer a common starting point for future research [21]. Therefore, we

considered the MIC value for the ODI-Chinese to be 11.2 point after 6-weeks of physical treatment.

In clinical practice, we should analyze the information (including the response to treatment for individual patients) in detail, rather than simply taking the reported MIC values when deciding whether a treatment is effective. Furthermore, comparison of all items in the instrument before and after treatment can help the clinician understand what items account for the change and adjust treatment program.

Limitation

A limitation of the study is that the sample size was low, especially when it was split into two groups (improved and non-improved). We should continue to enroll more patients and compare the responsiveness of subgroup for different degrees of disability and types of treatment.

Conclusions

The Chinese version of ODI was shown to be responsive in detecting clinical changes in subjects with chronic low back pain receiving conservative therapy.

Acknowledgments No benefits of any form have been received, or will be received from a commercial party related either directly or indirectly to the subjects of this article. No funds were received in support of this work.

References

1. Beaton DE (2000) Understanding the relevance of measured change through studies of responsiveness. *Spine* 25:3192–3199 (PMID: 11124736)
2. Beurskens AJ, de Vet HC, Köke AJ (1996) Responsiveness of functional status in low back pain: a comparison of different instruments. *Pain* 65:71–76 (PMID: 8826492)
3. Coelho RA, Siqueira FB, Ferreira PH, Ferreira ML (2008) Responsiveness of the Brazilian-Portuguese version of the Oswestry Disability Index in subjects with low back pain. *Eur Spine J* 17:1101–1106. (PMID: 18512083). doi:10.1007/s00586-008-0690-1
4. Costa LO, Maher CG, Latimer J (2007) Self-report outcome measures for low back pain: searching for international cross-cultural adaptations. *Spine* 32:1028–1037 (PMID: 17450079)
5. Davidson M, Keating JL (2002) A comparison of five low back disability questionnaires: reliability and responsiveness. *Phys Ther* 82:8–24 (PMID: 11784274)
6. Deyo RA, Centor RM (1986) Assessing the responsiveness of functional scales to clinical change: an analogy to diagnostic test performance. *J Chronic Dis* 39:897–906 (PMID: 2947907)
7. Fairbank JC, Couper J, Davies JB, O'Brien JP (1980) The Oswestry low back pain questionnaire. *Physiotherapy* 66:271–273 (PMID: 6450426)
8. Fairbank JC, Pynsent PB (2000) The Oswestry Disability Index. *Spine* 25:2940–2952 (PMID: 11074683)
9. Frost H, Lamb SE, Stewart-Brown S (2008) Responsiveness of a patient specific outcome measure compared with the Oswestry Disability Index v2.1 and Roland and Morris Disability Questionnaire for patients with subacute and chronic low back pain. *Spine* 33:2450–2458 (PMID: 18824951)
10. Grotle M, Brox JJ, Vøllestad NK (2004) Concurrent comparison of responsiveness in pain and functional status measurements used for patients with low back pain. *Spine* 29:E492–E501 (PMID: 15507789)
11. Guyatt GH, Berman LB, Townsend M, Taylor DW (1985) Should study subjects see their previous responses? *J Chronic Dis* 38:1003–1007 (PMID: 4066888)
12. Hashimoto H, Komagata M, Nakai O, Morishita M, Tokuhashi Y, Sano S, Nohara Y, Okajima Y (2006) Discriminative validity and responsiveness of the Oswestry Disability Index among Japanese outpatients with lumbar conditions. *Eur Spine J* 15:1645–1650. doi:10.1007/s00586-005-0022-7
13. Hägg O, Fritzell P, Nordwall A (2003) The clinical importance of changes in outcome scores after treatment for chronic low back pain. *Eur Spine J* 12:12–20. doi:10.1007/s00586-002-0464-0
14. Jin K, Sorock GS, Courtney T, Liang Y, Yao Z, Matz S, Ge L (2000) Risk factors for work-related low back pain in the People's Republic of China. *Int J Occup Environ Health* 6:26–33 (PMID: 10637534)
15. Lauridsen HH, Hartvigsen J, Manniche C, Korsholm L, Grønnet-Nilsson N (2006) Responsiveness and minimal clinically important difference for pain and disability instruments in low back pain patients. *BMC Musculoskelet Disord* 7:82. doi:10.1186/1471-2474-7-82
16. Liu H, Tao H, Luo Z (2009) Validation of the simplified Chinese version of the Oswestry Disability Index. *Spine* 34:1211–1217 (PMID: 19444069)
17. Li L, Wang HM, Shen Y (2003) Chinese SF-36 Health Survey: translation, cultural adaptation, validation, and normalization. *J Epidemiol Community Health* 57:259–263 (PMID: 12646540)
18. Mannion AF, Junge A, Grob D, Dvorak J, Fairbank JC (2006) Development of a German version of the Oswestry Disability Index. Part 2. Sensitivity to change after spinal surgery. *Eur Spine J* 15:66–73. doi:10.1007/s00586-004-0816-z
19. Merkesdal S, Mau W (2005) Prediction of costs-of-illness in patients with low back pain undergoing orthopedic outpatient rehabilitation. *Int J Rehabil Res* 28:119–126 (PMID: 15900181)
20. Norman GR, Stratford P, Regehr G (1997) Methodological problems in the retrospective computation of responsiveness to change: the lesson of Cronbach. *J Clin Epidemiol* 50:869–879 (PMID: 9291871)
21. Ostelo RW, Deyo RA, Stratford P, Waddell G, Croft P, Von Korf M, Bouter LM, de Vet HC (2008) Interpreting change scores for pain and functional status in low back pain: towards international consensus regarding minimal important change. *Spine* 33:90–94 (PMID: 18165753)
22. Taylor SJ, Taylor AE, Foy MA, Fogg AJ (1999) Responsiveness of common outcome measures for patients with low back pain. *Spine* 24:1805–1812 (PMID: 10488511)
23. Vianin M (2008) Psychometric properties and clinical usefulness of the Oswestry Disability Index. *J Chiropr Med* 7:161–163 (PMID: 19646379)
24. Walsh TL, Hanscom B, Lurie JD, Weinstein JN (2003) Is a condition-specific instrument for patients with low back pain/leg symptoms really necessary? The responsiveness of the Oswestry Disability Index, MODEMS, and the SF-36. *Spine* 28:607–615 (PMID: 12642770)
25. Wells G, Beaton D, Shea B, Boers M, Simon L, Strand V, Brooks P, Tugwell P (2001) Minimal clinically important differences: review of methods. *J Rheumatol* 28:406–412 (PMID:11246688)