In: Low Back Pain: New Research
Editor: M. C. Jansson and W. J. Lindberg, pp.
ISBN 978-1-60456-532-4

Chapter 6

Tools for the Evaluation of Low Back Pain Impairment: A Critical Review

Marco B. L. Rocchi*, Davide Sisti, Annarita Calavalle and Vilberto Stocchi
Istituto di Biomatematica, Università di Urbino "Carlo Bo"and
Istituto di Ricerca per l'Attività Motoria,
Università di Urbino "Carlo Bo"

Abstract

The aim of this study was to critically compare twelve self-administered questionnaires designed to evaluate disability caused by back pain. The questionnaires were analyzed considering and comparing their psychometric characteristics (reliability, validity, responsiveness, minimal clinically relevant difference), together with other practical and technical aspects (number of items, number and kind of domains, scaling of items, scoring, time to complete, etc). Data were obtained from scientific literature. Only 4 out of 12 analyzed instruments (i.e., the Oswestry Low Back Pain Disability Questionnaire, the Quebec Back Pain Disability Scale, Bournemouth Questionnaire for Measuring Outcome in Patients with Low Back Pain, and the Roland-Morris Low Back Pain Disability Questionnaire) appear fully validated from a psychometric standpoint.

On the basis of psychometric evaluations as well as feasibility considerations, the authors suggest using either the Roland-Morris or Oswestry questionnaire as the best assessment of the level of disability caused by back pain.

Introduction

The evaluation of disability caused by Low Back Pain (LBP) has become an important issue in the last few years. It is important not only as a tool to quantify the functional

* Tel: +39-0722-304256, Fax: +39-0722-304269, E-mail: marco.rocchi@uniurb.it

limitation of patients, but also as a way of assessing the results of treatment of people with LBP.

Hence, in the last few years several evaluative tools have been proposed in literature for this purpose. However, although they are published in authoritative scientific journals, not all these instruments appear completely validated from a psychometric standpoint.

Moreover, several papers have been published in the last years regarding a comparison among some of the considered questionnaires. Nevertheless, most of these comparisons were focused on their clinical uses more than on their psychometric properties [1-5]. Beurskens et al in 1995 [1] underlined the needing of addictional research in order to compare the existing questionnaires.

In this review we analyze twelve different questionnaires for the assessment of LBP impairment with the aim of providing useful psychometric-based information to those who are choosing among these questionnaires.

The Questionnaires

In this critical comparative review we considered twelve widely used self-administered questionnaires for the evaluation of disability in patients suffering from LBP:

- Oswestry Low Back Pain Questionnaire [6]
- Quebec Back Pain Disability Questionnaire [7-8]
- Bournemouth Questionnaire for Measuring Outcome in Patients with Low Back Pain [9]
- Aberdeen Low Back Pain Scale [10]
- Low-Back Outcome Scale by Greenough and Fraser [11]
- Back Pain Function Scale by Stratford et al. [12]
- Chronic Disability Index for Patients with Low Back Pain by Waddel and Main [13]
- Roland-Morris Disability Questionnaire [14]
- Modification of the Von Korff Pain-Scales to Evaluate Patients with Back Pain [15]
- Pain Disability Index [16]
- Dallas Pain Questionnaire [17]
- Million Disability Questionnaire [18]

We choose these questionnaires after a computer aided search from scientific literature, consulting the more relevant databases (such as: PubMed, EmBase, The Medical Algorithm Project).

In this review, we did not consider other widely used instruments, such as: the SF-36 [19] and some algometric questionnaires [20]. These evaluative tools were not included because of their lack of disease-specificity. Moreover, some disease-specific tools were excluded from this comparison because they are not self-administered questionnaires, e.g. the Functional Improvement Measures for Low Back Pain [21], the Fear-Avoidance Beliefs Questionnaire for Patients with Back Pain [22], the Low Back Pain Impairment Score [23],

the Clinical Symptoms Score of the Japanese Orthopaedic Association for a Patient with Lumbar Disc Herniation [24].

All the considered questionnaires were originally written and validated in English language. Later, some of them have been translated and re-validated in other languages.

For each questionnaire examined, we analyzed several factors that should determine its quality.

First of all, we consider the psychometric validation (i.e., the assessment of reliability, validity, responsiveness, and the definition of the minimal clinically relevant difference) [25].

Let us briefly review the meanings of these psychometric characteristics.

Reliability

The reliability of a measurement instrument is the extent to which it yields consistent, repeatable and reproducible estimates of what is assumed to be an underlying true score.

Three different approaches are usually used to evaluate reliability [25-26]:

a) Test-retest approach: in this approach the capability of the questionnaire to provide the same score on two consecutive occasions is assessed assuming that the disability of the patient to whom it is applied has not changed in the intervening period. A quantitative measure is provided by the Intra Class Coefficient (ICC).
b) Internal consistency approach: in this approach the degree of homogeneity of the items in a questionnaire is evaluated. This can be done by evaluating the correlation between each item of the questionnaire and the score of the whole questionnaire (usually excluding the item in question). Cronbach's alpha (i.e., the item-total correlation, reflecting the strength of the relationship between a single question and the total score) is used for this purpose.
c) Split-halves approach: in this approach, the items of the questionnaire are randomly split in two halves and the correlation of scores derived from each half is calculated.

Validity

The validity of a measurement questionnaire is determined according to whether particular interpretations of its scores are well justified. In other words, a measurement tool can be considered valid if it is really able to measure what it aims to measure.

Usually, given the lack of a universally accepted Gold Standard for assessing the degree of disability, several different types of validity are taken into consideration [27-28]:

a) Face validity: it is the validity assessed by those who might use it or those on whom it might be used. Their judgment is based on whether the content of the instrument appears to be relevant to the construct (e.g., disability) to be measured. It is also defined as credibility.

b) Content validity: in this approach it is evaluated whether the items of the questionnaires are linked by a plausible and explicit rationale to some particular conception of the construct to be measured (e.g., disability). This kind of validity should be assessed by experts and potential users. It is also defined as comprehensiveness.
c) Criterion-related validity: in this case, the validity can be evaluated in two different ways: by comparison with other instruments already validated (concurrent validity, measured by a correlation coefficient), or by the capability of the questionnaire to predict a certain event (predictive validity, evaluated by a regression model).

Responsiveness

The responsiveness of a measurement questionnaire refers to the magnitude of the change in scores associated with a given change in disability over time. It is also defined as sensitivity to change. From a statistical standpoint, it is usually evaluated by different methods [29-32]: the Area Under the Curve (AUC) of the ROC curves, the effect sizes, and the Guyatt's Responsiveness Index (GRI).

Nevertheless, as pointed out by Sheiner and Norman [25], there is no consensus regarding the appropriated measure of sensitivity of a measure to the main effect of treatment.

Minimal Clinically Relevant Difference

The minimal clinically relevant difference is defined as the smallest difference that clinicians (and/or patients) would care about.

Jaeschke et al [33] defined the MCRD as the smallest difference in score in the domain of interest which patients perceive as beneficial and which would mandate, in the absence of troublesome side-effects and excessive cost, a change in the patient's management.

This psychometric characteristic assumes a particular relevance in defining the sample size in clinical trials. It is usually calculated from the ROC curve using the cutoff point nearest the upper left-hand corner of the graph.

In this review, together with these psychometric properties, we also took into consideration, if available, other relevant aspects, such as: purpose and rationale, time to complete, number of items, domains and categories, scaling of items, final score range and its interpretation, field of application and clinical use.

All information was obtained from scientific literature.

Results

The results of this critical review are represented by the critical evaluation of the twelve analyzed questionnaires. The more relevant psychometric findings are summarized in Table I.

Table I. Summary Table of Psychometric Properties

Questionnaire	Reliability	Validity	Responsiveness	Minimal Clinically Relevant Difference
Oswestry	*	*	*	*
Quebec	*	*	*	*
Bournemouth	*	*	*	*
Aberdeen	*	*	*	
Greenough-Fraser	*	*		
Stratford	*	*		
Waddel-Main	*	*		
Roland-Morris	*	*	*	*
Von Korff	*	*		
Pain	*	(*)		
Dallas	*	(*)		
Million	(*)	(*)		

Asterisks indicate the verification of psychometric properties.
Asterisks within brackets indicate a partial verification of psychometric properties.

Oswestry Low Back Pain Questionnaire – This questionnaire was developed by Fairbank and Davies at the Orthopaedic Hospital in Oswestry (Shropshire, England). Its purpose is to assess patients suffering from LBP. It is a self-administered questionnaire, composed by 10 items describing the impact of pain on different daily living activities (e.g., personal care, walking, sitting, standing, traveling, sexual activities, etc.). It takes about 5 minutes to complete. Each item is scaled on a 6-point Likert scale (range 0-5), with 0 indicating no limitation due to pain and 5 indicating the impossibility of performing the activity in question. The total score ranges from 0 to 50, but it is usually converted in percent of disability by doubling it. Interpretation is linked to a 20%-interval: from 0-20% indicating a minimal disability, to 81-100% indicating either bed-ridden patients or patients exaggerating their symptoms. The Oswestry questionnaire is used as both an assessment tool in making prognoses, and as a way of measuring outcomes in clinical trials. From a psychometric standpoint, its reliability has been verified by test-retest (ICC = 0.94) and internal consistency (Cronbach's alpha = 0.93). It has good content, face and criterion-related validity (concurrent validity: good correlation with Chronic Disability Index by Waddell and Main r = 0.70, and with Roland-Morris Disability Questionnaire r = 0.70). The responsiveness is fairly good (AUC = 0.94, GRI = 3.49), and the minimal clinically relevant difference is 6 points.

Quebec Back Pain Disability Questionnaire – This questionnaire was developed by Kopec et al. at different Hospitals in Montreal, Toronto and London. Its purpose is to measure functional disability in patients with back pain. It is a self-administered questionnaire consisting of 20 items that describe the perceived difficulty of performing simple physical activities. The 20 items cover 6 domains (bed/rest; sitting/standing, ambulation, movement, bending/stooping, handling large or heavy objects). It takes about 5-10 minutes to complete. Each item is scaled on a 6-point Likert scale (range 0-5), with 0 indicating no perceived difficulty due to back pain and 5 indicating the impossibility of performing the activity. The total score ranges from 0 to 100, and it can be interpreted as the

percent of perceived disability. The Quebec questionnaire is used as both a tool to establish the severity of the diagnosis and prognosis, and to measure outcomes in clinical trials. From a psychometric standpoint, its reliability has been verified by test-retest (ICC = 0.92) and internal consistency (Cronbach's alpha = 0.96). It has good content, face and criterion-related validity (concurrent validity: good correlation with Roland-Morris Disability Questionnaire r = 0.80). Its responsiveness is fairly good (AUC = 0.87, GRI = 1.82), and its minimal clinically relevant difference is 15 points.

Bournemouth Questionnaire for Measuring Outcome in Patients with Low Back Pain – This questionnaire was developed by Bolton and Breen at Anglo-European College of Chiropratic in Bournemouth (England). Its purpose is to provide an easy to use outcome measure in patients with LBP. It is a self-administered questionnaire, consisting of 7 items that aim to measure the interference caused by LBP in various domains (pain, emotion, daily activities) in the last week. It takes less than 5 minutes to complete. Each item is scaled on an 11-point Likert scale (range 0-10), with 0 indicating no interference and 10 indicating the maximum interference due to back pain. The total score ranges from 0 to 70. The higher the score the greater the impact of LBP on the patient's life. The Bournemouth questionnaire is used to establish the severity of the diagnosis and to make prognoses. From a psychometric standpoint, its reliability has been verified by test-retest (ICC = 0.95) and internal consistency (Cronbach's alpha = 0.90). It has acceptable content, face and criterion-related validity and responsiveness (effect sizes = 1.29). Its minimal clinically relevant difference has been evaluated in 5 points.

Aberdeen Low Back Pain Scale – This questionnaire was developed by Ruta et al. at the University of Aberdeen and Aberdeen Royal Infirmary (Scotland). Its purpose is to measure outcome in patients with LBP. It is a self-administered questionnaire consisting of 19 items that aim to measure the impact of pain on different activities (e.g., personal care, walking, sitting, standing, sexual activities, etc.). It takes about 20 minutes to complete. Each item is scaled differently (some on a 6-point, 5-point, 3-point Likert scale, others adding 1 point for a positive answer). The total score ranges from 0 to 75, but it is usually converted into percentage by a proportion. The higher the score, the greater the severity of the back pain. The Aberdeen questionnaire is essentially used for initial evaluation of patients and to monitor the effectiveness of therapeutic interventions; however, it has also recently been used to measure outcomes in clinical trials. From a psychometric standpoint, its reliability has been verified by internal consistency (Cronbach's alpha = 0.80). It has good content, face and criterion-related validity (concurrent validity: good correlation with Chronic Disability Index by Waddell and Main, and with Oswestry Low Back Pain Disability Questionnaire). The responsiveness is fairly good (effect size = 0.62). The minimal clinically relevant difference has not yet been evaluated.

Low-Back Outcome Scale by Greenough and Fraser – This questionnaire was developed by Greenough and Fraser at the Middleborough General Hospital (England) and Royal Adelaide Hospital (Australia). Its purpose is to measure functional outcome in patients with LBP. It is a self-administered questionnaire consisting of 13 items that measure the impact of LBP on several domains (e.g., pain, employment, sport and social activities, resting, sex life, etc.). It takes about 10 minutes to complete. Each item has 3 to 4 possible answers. A particular score is assigned to each answer, and each item has its own point value (e.g., 0-3-6-

9 for employment, 0-1-2-3 for walking and sitting, etc.). The total score ranges from 0 to 75. The higher the score the better the patient's status. Interpretation of the final score ranges from 0-29 points indicating a poor status, to 65-75 points indicating an excellent status. The Low Back Outcome Scale questionnaire is used both in making prognoses, and to measure outcomes in clinical trials. From a psychometric standpoint, its reliability has been verified by test-retest (ICC = 0.84) and internal consistency (Cronbach's alpha = 0.85). It has good content and face validity, but its concurrent validity compared to other questionnaire has not been demonstrated. Responsiveness and minimal clinically relevant difference have not been assessed.

Back Pain Function Scale by Stratford et al. – This questionnaire was developed by Stratford et al. at two different universities in Georgia and Virginia (USA). It aims to evaluate functional ability in patients suffering from back pain. It is a self-administered questionnaire consisting of 12 items that measure the patient's ability to perform simple physical activities. The 12 items cover different domains (social activities, sitting/standing, ambulation, movement, bending/stooping, handling large or heavy objects, etc). It takes about 5 minutes to complete. Each item is scaled on a 6-point Likert scale (range 0-5), with 0 indicating the inability to perform the action due to back pain and 5 indicating no difficulty. The total score ranges from 0 to 60, but the authors suggest that it be converted in a 0-1 scale. The higher the score the greater the patient's functional ability. The Back Pain Function Scale is mainly used in establishing the severity of a diagnosis. From a psychometric standpoint, its reliability has been verified by test-retest (ICC = 0.88) and internal consistency (Cronbach's alpha = 0.93). It has good content, face and criterion-related validity (predicted validity: r = 0.65). Responsiveness and minimal clinically relevant difference have not been assessed.

Chronic Disability Index for Patients with Low Back Pain by Waddel and Main – This questionnaire was developed by Waddel and Main at the Western Infirmary in Glasgow (Scotland). It aims to provide an easy-to-use disability index for patients with LBP. It is a self-administered questionnaire consisting of 9 items that aim to measure the patient's ability to perform simple daily activities. The 9 items cover different domains (social activities, sitting/standing, ambulation, sleeping, handling heavy objects, etc). It takes about 3-5 minutes to complete. Each item has a binary answer (yes/no). The total score is obtained by adding all the yes responses to give a score out of 9. The higher the score the higher the level of disability. The Chronic Disability Index is chiefly used as discriminative measure to assess the severity of LBP. From a psychometric standpoint, its reliability has been validated exclusively by internal consistency (Theta statistic = 0.76). It has good face and content validity and fair criterion-related validity (concurrent validity: discrete correlation with Oswestry Questionnaire r = 0.70). Responsiveness and minimal clinically relevant difference have not been established.

Roland-Morris Disability Questionnaire – This questionnaire was developed by Roland and Morris at the St.Thomas Hospital in London. It aims to evaluate the level of disability in patients with LBP. It is a self-administered questionnaire consisting of 24 items that measure the interference of LBP in different domains (e.g., mobility, dressing, working, standing, sleeping, mood, recreation, appetite). It takes about 5 minutes to complete. Each item has a binary answer (yes/no). The total score is obtained by totaling yes responses to give a score out of 24. The higher the score the more severe the disability caused by LBP. The Roland-

Morris questionnaire was designed to aid in making prognoses, but it has also been used to measure outcomes in clinical trials. From a psychometric standpoint, its reliability has been verified by test-retest (ICC = 0.91) and internal consistency (Cronbach's alpha = 0.93). It has good content, face and criterion-related validity (concurrent validity: good correlation with Oswestry questionnaire, r = 0.89). Its responsiveness is fairly good (AUC = 0.79), and it has a minimal clinically relevant difference of 4-5 points.

Modification of the Von Korff Pain-Scales to Evaluate Patients with Back Pain – This questionnaire was designed by Underwood et al at the St.Bartholomew's and the Royal London School of Medicine in London. It was developed by modifying the pain scales devised by Von Korff et al. Its purpose is to evaluate the level of pain in patients with LBP. It is a self-administered questionnaire consisting of 6 items that aim to measure the interference of pain in two different domains: disability (items 1-3) and pain (items 4-6). It takes about 3-5 minutes to complete. Each item is scaled on an 11-point Likert scale (range 0-10), with 0 indicating no interference and 10 indicating maximum interference caused by pain. The questionnaire gives two different scores for disability and pain. These total scores range from 0 to 30, but the authors suggest converting them in a 0-100 scale. The higher the score the more severe the disability or the back pain. The Modification of the Pain Scales of the Von Korff questionnaire was designed to measure outcomes in clinical trials. From a psychometric standpoint, its reliability has been verified by test-retest (ICC = 0.87). It has good content, face and criterion-related validity (concurrent validity: good correlation with Roland-Morris questionnaire, r = 0.87). Responsiveness and minimal clinically relevant difference have not been assessed.

Pain Disability Index – This questionnaire was developed by Tait and collegues at the St.Louis University School of Medicine in St.Louis (Missouri, USA). Its purpose is to assess pain-related disability in patients suffering from chronic pain; although it has been originally validated on patients with chronic back pain, it has used also for pain of different origin. It is a self-administered questionnaire, composed by 7 items describing the impact of pain on 7 domains (family responsibilities, recreation, social activities, occupation, sexual behavior, self care, life-support activities). It takes about 3 minutes to complete. Each item is scaled on a 11-point Likert scale (range 0-11), with 0 and 10 indicating no disability and total disability, respectively. The total score ranges from 0 to 70; the higher the score the greater the impact of LBP on patients life. The Pain Disability Index is used both for clinical monitoring and for research supports. From a psychometric standpoint, its reliability has been verified only by internal consistency (Cronbach's alpha = 0.87). It has good content and face validity but its criterion-related validity (concurrent and predictive validity) has not been evaluated. Responsiveness and minimal clinically relevant difference have not been assessed.

Dallas Pain Questionnaire – This questionnaire was developed by Lawlis and collegues at the Medical Arts Hospital in Dallas (Texas, USA). Its purpose is to assess the impact of spinal pain on behavior. It is a self-administered questionnaire, composed by 16 items describing the impact of pain on 4 domains (daily activities, work/leisure activities, anxiety/depression, social interests). It takes about 5 minutes to complete. Each item is associated to a visuo-analogue scale, ranging from 0 to 1. For each domain, the score is corrected with different weights. The total score ranges from 0 to 66; the higher the score the greater the impact of spinal pain on patients life. The Dallas Pain Questionnaire is essentially

used for clinical monitoring. From a psychometric standpoint, its reliability has been verified by both test-retest (ICC = 0.97) and internal consistency (item-total correlation, ranging from 0.44 to 0.94). It has good content, face and criterion-related validity only from a psychological standpoint (significant concurrent validity with the Minnesota Multiphasic Psychological Inventory). Responsiveness and minimal clinically relevant difference have not been assessed.

Million Disability Questionnaire – This questionnaire was developed by Million and collegues at the Hope Hospital of the University of Manchester (England). Its purpose is to assess the severity of symptoms in patients affected by LBP. It is a self-administered questionnaire, composed by 15 items subdivided into 2 domains (symptoms and their interference with daily activities). It takes about 4 minutes to complete. Each item is associated to a visuo-analogue scale, ranging from 0 to 1. The total score ranges from 0 to 15; the higher the score the greater the impact of LBP on patients life. The Million Disability Questionnaire is used for clinical monitoring. From a psychometric standpoint, its reliability has been verified by test-retest, but a correlation coefficient was used instead of the intraclass coefficient (r = 0.97). It has acceptable content and face validity but its criterion-related validity (concurrent and predictive validity) has not been evaluated. Responsiveness and minimal clinically relevant difference have not been assessed.

Conclusion

Our comparative analysis of the twelve questionnaires designed to assess disability level in patients with LBP allows us to draw the following conclusions:

a) Only 4 out of the 12 questionnaires have been fully validated from a psychometric standpoint (i.e., their reliability, validity, responsiveness have been proved, and their minimal clinically relevant difference has been defined): the Oswestry, the Quebec, the Bournemouth and the Roland-Morris questionnaires. All the other questionnaires are valid and reliable (although in some cases only partially), but their responsiveness and minimal clinically relevant difference have yet to be demonstrated. At the moment, this lack of information makes them unsuitable for application at least as outcome measures in clinical trials.
b) Other considerations regard the readability and feasibility of the questionnaires. From this standpoint, the Roland-Morris, the Bournemouth, and the Quebec questionnaires appear slightly more readable than the other fully validated questionnaires. On the other hand, the articulated structure of the answers for each item in the Oswestry, probably allows for a deeper investigation of Low Back Pain phenomenon.
c) Lastly, a large number of validated translated versions of these four questionnaires (especially of the Rolend-Morris) are available in scientific literature.

Basing on these results, we can conclude that the final choice of a suitable instrument for the evaluation of Low Back Pain should obviously be made on the basis of the properties that

are the most relevant to the clinician's purposes. From this point of view, our contribution should only be considered complementary with other relevant aspects, such as the clinical considerations and the international uses. Nevertheless, based on psychometric evaluations but also on feasibility considerations, we think it is reasonable to suggest either the Roland-Morris or the Oswestry questionnaires as the tools of choice in the assessment of the severity of disability caused to back pain. The Quebec and the Bournemouth questionnaires also resulted fully validated, but, at the moment, their limitated use in scientific literature makes them a secondary choice instruments for Low Back Pain evaluation.

However, if possible, we suggest the combinated use of more than one questionnaire, in order to avoid ceiling and floor effects.

References

[1] Beurskens AJ, de Vet HC, Koke AJ, van der Heijden GJ, Knipschild PG. Measuring the functional status of patients with low back pain. Assessment of the quality of four disease-specific questionnaires. *Spine* 1995; 20:1017-28.

[2] Kopec JA, Esdaile JM. Functional disability scales for back pain. *Spine* 1995; 20:1943-9.

[3] Deyo RA, Battie M, Beurskens AJ, Bombardier C, Croft P, Koes B, Malmivaara A, Roland M, Von Korff M, Waddell G. Outcome measures for low back pain research. A proposal for standardized use. *Spine* 1998; 23:2003-13.

[4] Bombardier C. Outcome assessments in the evaluation of treatment of spinal disorders: summary and general recommendations. *Spine* 2000; 25:3100-3.

[5] Fritz JM, Irrgeng JJ. A comparison of a modified Oswestry Low Back Pain Disability Questionnaire and the Quebec Back Pain Disability Scale. Phys Ther 2001; 81:776-88.

[6] Fairbank JCT, Davies JB. The Oswestry low back pain disability questionnaire. *Physiother* 1980; 66:271-3.

[7] Kopec JA, Esdaile JM, Abrahamowicz M, Abenhaim L, Wood-Dauphinee S, Lamping DL, Williams JI. The Quebec back pain disability scale. Measurement properties. *Spine* 1995; 20:341-52.

[8] Kopec JA, Esdaile JM, Abrahamowicz M, Abenhaim L, Wood-Dauphinee S, Lamping DL, Williams JI. The Quebec back pain disability scale: conceptualization and development. *J. Clin. Epidemiol*. 1996; 49:151-61.

[9] Bolton JE, Breen AC. The Bournemouth Questionnaire: A short-form comprehensive outcome measure. I. Psychometric properties in back pain patients. *J. Manipul. Physiol*. Ther 1999; 22:503-10.

[10] Ruta DA, Garratt AM, Wardlaw D, Russell IT. Developing a valid and reliable measure of health outcome for patients with low back pain. *Spine* 1994; 19:1887-96.

[11] Greenough CG, Fraser RD. Assessment of outcome in patients with low-back pain. *Spine* 1992; 17:36-41.

[12] Stratford PW, Binkley JM, Riddle DL. Development and initial validation of the Back Pain Functional Scale. *Spine* 2000; 25:2095-102.

[13] Waddell G, Main CJ. Assessment of severity in low-back disorders. *Spine* 1984; 9:204-8.

[14] Roland M, Morris R. A study of the natural history of low-back pain. Part I: Development of a reliable and sensitive measure of disability in low-back pain. *Spine* 1983; 8:141-4.

[15] Underwood MR, Barnett AG, Vickers MR. Evaluation of two time-specific back pain outcome measures. *Spine* 1999;24:1104-12.

[16] Tait RC, Pollard A, Margolis RB, Duckro PN, Krause SJ. The Pain Disability Index: psychometric and validity data. *Arch. Phys. Med. Rehabil*. 1987; 68:438-41.

[17] Lawlis GF, Cuencas R, Selby D, McCoy CE. The development of the Dallas Pain Questionnaire. An assessment of the impact of spinal pain on behavior. *Spine* 1989; 14:510-6.

[18] Million R, Nilsen K, Jayson MI,Baker RD. Evaluation of lower back pain and assessment of lumber corsets with and without back supports. *Ann. Rheum. Dis*. 1981;40:449-54.

[19] Ware JE, Sherbourne CD. The MOS 36-Item Short Form Health Survey (SF-36). I: Conceptual framework and item selection. *Med. Care* 1992; 30:473-83.

[20] Price DD. The validation of visual analogue scale as ratio scale measures for chronic and experimental pain. *Pain* 1983;17:45-56.

[21] Gottlieb H, Strite LC, Koller R, Madorsky A, Hockersmith V, Kleeman M, Wagner J. Comprehensive rehabilitation of patients having chronic low back pain. *Arch. Phys. Med. Rehabil*. 1977; 58:101-8.

[22] Waddell G, Newton M, Henderson I, Somerville D, Main CJ. A Fear-Avoidance Beliefs Questionnaire (FABQ) and the role of fear-avoidance beliefs in chronic low back pain and disability. *Pain* 1993; 52:157-68.

[23] Waddell G, Main CJ. Assessment of severity in low-back disorders. *Spine*. 1984; 9: 204-208.

[24] Japanese Orthopaedic Association (JOA). Japanese Orthopaedic Association assessment criteria guidelines manual. Tokio: *JOA*; 1996.

[25] Streiner DL, Norman GR. Health measurement scales. A practical guide to their development and use. *Second edition*. Oxford: Oxford University Press; 1995.

[26] Shrout PE, Fleiss JL. Intraclass correlations: using in assessing rater reliability. *Psychol. Bull* 1979; 86:420-6.

[27] Deyo RA. Measuring functional outcomes in therapeutic trials for chronic disease. *Control Clin. Trials* 1984; 5:223-40.

[28] Kirshner B, Guyatt C. A methodological framework for assessing health indices. *J. Chronic. Dis*. 1985; 38:27-36.

[29] Deyo RA, Centor RM. Assessing the responsiveness of functional scales to clinical change: an analogy to diagnostic test performance. J Chronic Dis 1986; 39:897-906.

[30] Cohen J. Statistical power analysis for the behavioural sciences. New York: Academic Press; 1977.

[31] Guyatt GH, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. *J. Chronic. Dis*. 1987; 40:171-8.

[32] Beurskens AJ, de Vet HC, Hoke AJ. Responsiveness of functional status in low back pain: a comparison of different instruments. *Pain* 1996; 65:71-6.

[33] Jaeschke R, Singer J, Guyatt GH. Ascertaining the minimal clinically important difference. *Control Clin. Trials* 1989; 10:407-15.