Minimal Clinically Important Difference. Low Back Pain: Outcome Measures

CLAIRE BOMBARDIER, JILL HAYDEN, and DORCAS E. BEATON

ABSTRACT. A proposed standard "core set" of outcome measures for low back pain includes 5 domains: back-specific function, generic health status, pain, work disability, and patient satisfaction. This paper focuses on the 2 recommended back-specific measures of function: the Roland-Morris Disability Questionnaire (RDQ) and the Oswestry Disability Index (ODI). We specifically address their ability to measure change. A systematic review of the literature identified a total of 78 and 71 (RDQ and ODI, respectively) articles as potentially relevant. Detailed tables are provided for each citation, with the type of back pain population studied, the type of change measured, the estimate of change, and the interval over which the change was studied. These tables should be used as a reference for sample size calculation. The responsiveness of the RDQ found in the literature ranges from 2 to 8 points on its 0 to 24 scale depending on what change is being measured. As a rough guide, Roland recommends that a change in 2–3 points on the RDQ should be considered the minimum clinically important change. Choosing any value larger than 5 in designing a clinical trial would risk underpowering the trial, since fewer patients are needed if a trial is designed on the basis of a large change score. (J Rheumatol 2001;28:431–8)

Key Indexing Terms:
LOW BACK PAIN
OUTCOME ASSESSMENT
RANDOMIZED CLINICAL TRIALS
REPRODUCIBILITY OF RESULTS

INTRODUCTION

Clinical studies of low back pain (LBP) vary widely in the type of outcomes used to assess the success of therapies. There is no consistency in the range of outcomes considered and even within a given outcome the measures selected vary considerably. For example, there are more than 10 different tools to measure back-specific function, and pain scales differ from one study to another. A standard set of outcome measures would make it easier to compare the magnitude of the treatment effects across studies. It would also encourage complete reporting of relevant outcomes so that investigators do not report on a single domain while ignoring others.

For all these reasons, an international group of investigators met and proposed a core set of 5 domains that should be used in all low back pain studies. These recommendations were published in *Spine* in 1998[1]. Individual investigators may choose to add other measures in their studies, but the core set is intended to provide a common yardstick across all studies. The core set for clinical research is illustrated in Table 1. This core set has since been revisited for a forthcoming special issue of *Spine*[2a] and the same 5 domains are recommended with only slight differences in the specific instruments proposed for each domain. For example, the full Medical Outcome Survey Short Form-36 (SF-36) is now recommended instead of the SF-12 because of its superior measurement properties, particularly for detecting change in a clinical context.

In this paper, we focus on the 2 back-specific measures of function recommended in the "core-set," the Roland-Morris Disability Questionnaire (RDQ)[2b] and the Oswestry Disability Index (ODI)[3]. They are the most commonly used measures of function in back pain research, but how well do they detect patients' improvement and what is their smallest clinically relevant change? To answer these questions, we review the literature, identify the different approaches used to measure change, and document the magnitude of the change found. Finally, we classify our results using the framework proposed by Beaton elsewhere in this supplement[4a].

METHODS

The Roland-Morris Disability Questionnaire (RDQ) measures 24 activity limitations due to back pain, while the Oswestry Disability Index (ODI) consists of 10 items assessing the level of pain interference with physical activities; the RDQ is mostly a measure of function while the ODI incorporates a measure of pain as well as physical function. In practice, the differences between these instruments are

From the Institute for Work and Health, and Clinical Epidemiology and Health Research Program, Department of Medicine and Department of Public Health, University of Toronto, Toronto, Ontario, Canada.

C. Bombardier, MD, FRCP, Institute for Work and Health, Clinical Epidemiology and Health Care Research Program, Department of Health Administration, University of Toronto; J. Hayden, BSc, DC, Institute for Work and Health; D.E. Beaton, BScOT, MSc, PhD, Institute for Work and Health, Clinical Epidemiology and Health Care Research Program, Department of Health Administration, University of Toronto.

Address reprint requests to Dr. C. Bombardier, Institute for Work & Health, 250 Bloor St East, Ste 702, Toronto, Ontario M4W 1E6. E-mail: claire.bombardier@utoronto.ca

Table 1. A proposed "core set" of instruments for clinical researchers. Reproduced with permission from *Spine*[1].

Domain	Specific Instrument
Pain symptoms	Bothersomeness or severity and frequency of low back pain and leg pain (sciatica)
Back related function	Roland-Morris Disability Questionnaire (or adaptations) or Oswestry Disability Index (or adaptations)
Generic well being	SF-12 or EuroQol:also: "If you had to spend the rest of your life with the symptoms you have right now, how would you feel about it?"
Disability (social role)	Days of work absenteeism, cut down activities, bed rest
Satisfaction with care	Single question on overall satisfaction (optional)

Table 2. Comparison of the Roland-Morris Disability Questionnaire and the Oswestry Disability Index.

Instrument	No. of Items (response options)	Score (best to worse)	Time to complete, min	Dimensions
Roland-Morris	24 (yes/no)	0–24	5	Physical activities, housework, mobility, dressing, getting help, appetite, irritability, pain
Oswestry	10 (6 levels)	0–100	5	Pain intensity, personal care, lifting, walking, sitting, standing, sleeping, sex life, social life, traveling.

not great. They are both easy to use, reliable, and valid. Some floor and ceiling effects have been described; the ODI may be better at measuring higher disability levels, while the RDQ may perform better in general populations, where lower disability levels are found.

To assess their ability to measure change we undertook electronic and manual citation searches of the original publications of Roland and Morris (1983)[2b] and Fairbank, *et al* (1980)[3]. The electronic search covered the time period 1993 to March 2000; it identified 205 and 334 articles for the RDQ and ODI, respectively. Manual citation searches for the years prior to 1993, back to the year of publication of the instrument, identified a further 36 and 47 articles resulting in a total of 241 and 381 references for the RDQ and ODI, respectively. These articles were assessed by title and abstract (if available) to determine eligibility. A total of 78 and 71 (RDQ and ODI, respectively) articles were considered potentially relevant to responsiveness and were retrieved for closer examination. (Note that all articles identified from the manual search were retrieved since electronic titles and abstracts were unavailable for screening.)

The detailed results of our literature search are presented in Appendix 1 for the RDQ (3 articles could not be retrieved) and in Appendix 2 for the ODI (retrieval is continuing for this measure). These appendices offer a summary of this literature to date. For each citation, we extracted the type of back pain population studied, the type of change measured, the estimate of change, and the interval over which the change was studied. We classify the type of change measured using the framework proposed by Beaton[4a]: does the change refer to individuals or groups of individuals, does it describe a difference between individuals or within individuals or both, and what type of change is being studied?: minimum potentially detectable, minimum actually detectable beyond error, change observed in a population, or change observed in those estimated to have an *important difference*.

RESULTS

Changes in individuals. Changes in individuals were measured less frequently in the literature than changes in groups. Stratford, *et al*[4b] estimate that the minimally detectable change on the RDQ is 5 points on the 0–14 point scale, with less confidence of detecting changes at the extremes of the scale (Figure 1, number 1). Dionne, *et al*[5] estimate that a 10% change is the best discriminator between those who return to work and those who do not, or a change of 2 to 3 points on the scale (Figure 1, number 2). Stratford, *et al*[6] and Riddle, *et al*[7] calculate that a 5 point change is an important difference since it is the change found in patients where both the patient and the clinician believe an important change has occurred and treatment goals have been achieved (Figure 1, number 3). The approach most often used for deciding on an important change in individuals is to contrast the distributions of changes in the RDQ and in the ODI in individuals who meet a given criterion with the distribution of changes in individuals who do not meet that criterion. For example, Deyo and Centor[8] compared the distributions of changes in RDQ scores in those who resumed full activity to those who did

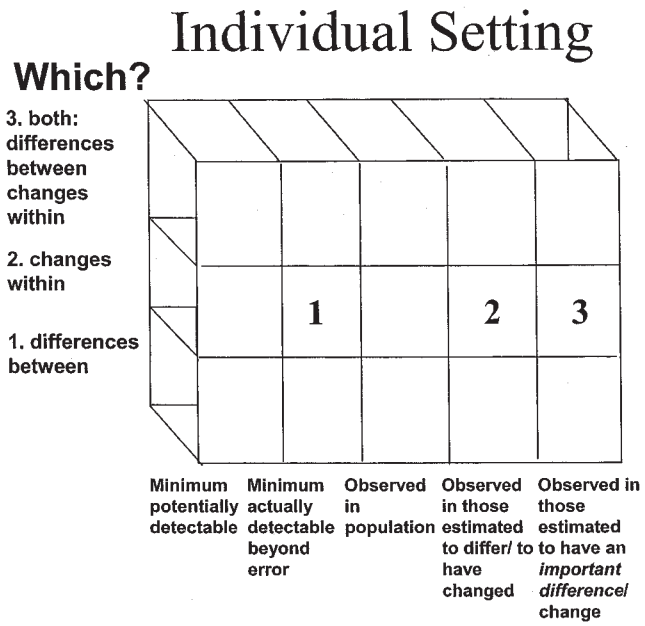

Figure 1. Examples of changes found in studies looking at individual patients. 1. Change beyond error > 5 Roland-Morris points, slightly less (4) at extremes of score. 2. Change of 2-3 is most accurate in discriminating improved versus nonimproved (self-rating). Ten percent change is the best discriminator between those returned to work and those not. 3. Clinician and patient rating of change, change of 5 points = most accurate indicator of improvement. The same is found when using achievement of treatment goals as indicator.

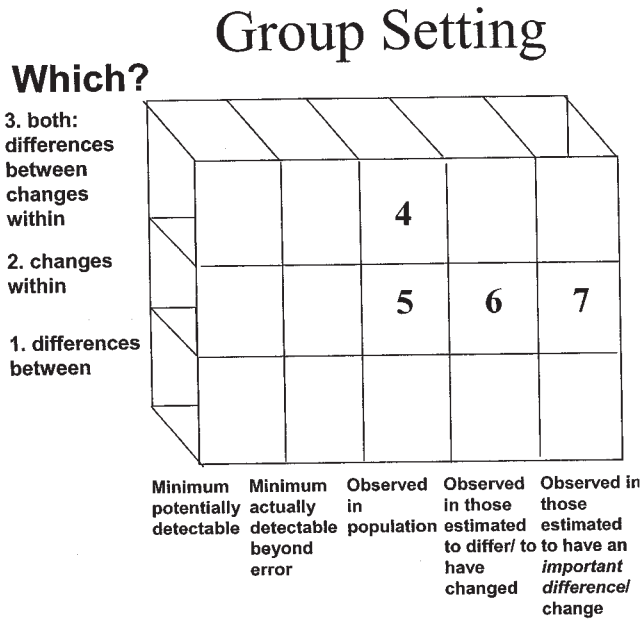

Figure 2. Examples of changes found in studies looking at group changes. 4. Randomized controlled trials using the Roland scale. Change depends on treatment, control. 5. Short term history of acute back pain, mean change = 3. Pre-post therapy = 6.2. 6. Mean change of 8.7 in patients with improved sciatica, improved quality of life (8.4). 7. Mean change in those patients deemed to have had an important improvement: 7.2–7.6: achieving treatment goals, clinician/patient rating.

not resume full activity. Beurskens, *et al*[9] contrasted the RDQ change scores in those who improved using a global assessment of improvement versus those who did not, while Stratford, *et al*[6,10] used as a criterion improvements according to both the clinician and the patients' global rating. All these authors used a well known method called the receiver operand curve (ROC), which results in various cutoff points for changes, each associated with a given sensitivity and specificity.

Changes in groups. There are many studies looking at changes in groups of individuals. The most common ones are randomized trials where the treatment effects for the 2 back-specific measures, the RDQ or the ODI, can be calculated. The treatment effect is the mean change in the treatment group contrasted with the mean change in the comparison group (Figure 2, number 4); more specifically, it is the difference in observed changes between the 2 cohorts.

Appendices 1 and 2 provide examples of treatment effects in randomized trials[11-16]. Another type of change is illustrated in Figure 2, number 5; it is the change found in a single cohort of patients. The magnitude of this change will vary depending on the patient population (acute vs chronic back pain), on the length of followup, or on the types of interventions used during the followup period (surgery vs analgesics). Several examples of cohorts are included in the appendices[8,10,17-19]. A third type of change shown in Figure 2, number 6, is the change observed in those estimated to have changed. Patrick, *et al*[19] found a change of 8.7 for those with improved sciatica and 8.4 for those who had an improvement in quality of life. The final example in Figure 2, number 7, relates to the change observed in those estimated to have an important change. Achieving treatment goals is deemed an important change — Riddle, *et al*[7] found that the magnitude of such a change varies depending on the baseline value; Stratford, *et al*[6] found the same for the average of the clinician and the patients' global rating of important change; but on average, the change was in the range of 7 points.

DISCUSSION

The most common reason for using patient based outcome measures is to assess patients' response to treatment. Is the patient better? This paper reviewed the types and magnitude of changes found in 2 measures of back-specific function: the Roland-Morris Disability Questionnaire and the

Oswestry Disability Index. Both are part of the recommended "core set" of measures for use in back pain research. For a comprehensive review of other measurement properties of these 2 scales the reader can refer to a recent review by Roland and Fairbank[2a].

We found that there is no set answer to the question of what is an important change. The magnitude of change varies depending on the approach used to measure change. Some authors label as responsiveness any change greater than the measurement error of the instrument, others use the change observed in a cohort of patients, while others restrict it to the change found in patients who said they improved. It is not surprising, then, that the responsiveness of the RDQ found in the literature will range from 2 to 8 points on its 0 to 24 scale depending on what change is being measured. Since the decision about what is a clinically important change is key for sample size calculations, the reader should carefully apply Beaton's guidelines and refer to the tables in the appendices of this paper to find a match between the responsiveness reported in the literature and the study you are planning. As a rough guide Roland recommends that a change in 2 to 3 points on the RDQ should be considered the minimum clinically important change. Choosing any value larger than 5 in designing a clinical trial would risk underpowering the trial, since fewer patients are needed if a trial is designed on the basis of a large change score.

Appendix 1. Roland-Morris Disability Questionnaire.

Author	Patient Population	Which comparison	Who is focus	Type of change	Estimate of change (/24)	Interval
Lanier & Stockton, 1988(17)	LBP <28 days; presentation to GP	Over time	Group	Observed	Mean: 6-week: 9.1	6 weeks
Deyo & Centor, 1986(8) (24 items of RMDQ from SIP)	LBP; presentation to hospital walk-in clinic; n=120	Over time	A. Group B. Individual	Observed: 1. Natural history of LBP Estimated: 2. Resumed full activities 3. Patient and examiner: improved	A. Mean (sd) 1. 3.0 (5.2) 2. 3.8 (5.3) 3. 4.4 (5.4) B. ROC area (se) 2. 0.72 (.047) 3. 0.67 (.068)	3-week
Beurskens et al., 1996(9)	Non-specific LBP >6 weeks; n=81; treatment = traction.	Over time	A. Group B. Individual	Estimated: Global perceived effect	A. Mean (sd): 7.8 (3.9) Effect size: 2.02 B. ROC area: 0.93 (change 2.5-5 points)	5 weeks
Patrick et al., 1995(19) (**Modified 23-item)	Sciatica; presenting to orthopedic and neurologic surgeon offices; surgical vs. non-surgical treatment; n=427	Over time	Group	Observed: 1. Treatment effect Estimated: 2. Improved leg pain (patient) 3. Improved quality of life (patient)	A. Mean : 1. 7.2 2. 8.7 3. 8.4 Effect size: 2. 1.6 3. 1.6	3 months
Stratford et al., 1996(4)	Musculoskeletal LBP; referred to outpatient physiotherapy; n=60; Physiotherapy treatment	Over time	Individual	Minimum detectable change (CSEM)	Mean (90% CI): 5 (4, 5) *cannot confidently detect improvement in patients with initial scores <4, or deterioration if >20.	4 to 6 weeks
Constant et al., 1995(13)	Chronic LBP (>1 year); GP referred; n=62; Spa therapy treatment	Between group, within person	Group	Observed (Treatment, control groups)	Mean (sd) 3 week: Tx - 3.5 (4.4) C - 0.1 (2.7) 6 month: Tx - 5.1 (4.4) C - 0.9 (3.4)	3 weeks, 6 months

Author	Patient Population	Which comparison	Who is focus	Type of change	Estimate of change (/24)	Interval
Burton et al., 1999(11)	Non-specific LBP; new episode of acute or recurrent (<3 months); primary care (GP, osteopath) practices; n=162; Educational booklet	Over time	Group	Observed: Treatment effect	No significant differences between groups; NB. Study assessed RR of "Important" change in two groups as: a. >4 point improvement in FABphys b. >3 point improvement in RMDQ	2 week, 3 month, 1 year
Sinclair et al., 1997(20)	Soft tissue injuries inception cohort; WCB claimants; n (LBP)=885; Community clinic program	Between group, within person	Group	Observed: Treatment effect	No significant differences in rate of change between groups	4, 10, 16, 52 weeks
Leclaire et al., 1996(15)	Work related LBP <3 months; Private physiatric outpatient clinic attendants; n=168; physiotherapy treatment or physio + back school	Over time	Groups	Observed: Treatment effect	Difference between groups (95% CI): Post-treatment 6.54 (1.59, 11.49) No significant differences at 6, 12 months	Post-treatment, 6, 12 months
Cherkin et al., 1998(12) (**Modified 23-item)	LBP >7 days following primary care presentation; n=323; physiotherapy, chiropractic, educational booklet	Over time	Group	Observed: a. booklet b. chiropractic c. physiotherapy	Mean: 4 weeks a. 6.8 b. 8.4 c. 8.1 12 weeks a. 7.4 b. 9 c. 8.1	4, 12 weeks
Constant et al., 1998(14)	Chronic LBP (>1 year); GP referred; n=224; Spa therapy treatment added to usual drug treatment	Between group, within person	Group	Observed: a. drug b. drug + spa	Mean (sd): 3 weeks a. 0.6 (3.4) b. 3.4 (3.4) 3 months a. 1.1 (3.1) b. 4.0 (4.1)	3 weeks, 3 months
Dionne et al., 1999 (5) (**Modified 16-item)	LBP; GP visit, HMO enrollees; n=720; standard treatment	Over time	Individual	Estimated: Work status improvement at 2 years (employed vs. unemployed)	ROC analysis: 10% decrease maximum discrimination (sens=.84, spec=.54)	2 years
Stratford et al., 1994(10)	Mechanical LPB; referred by GP to physiotherapy; n=88;	Over time	A. Group B. Individual	Observed: 1. Treatment effect Important: 2. Average of clinician and patient global rating, importance of change (>5 on 15 point scale)	Mean (sd): 1. 4.7 (5.0) Correlation with criterion change: 2. 0.60 ROC analysis: 2. ROC area=0.79	4 to 6 weeks

Author	Patient Population	Which comparison	Who is focus	Type of change	Estimate of change (/24)	Interval
Stratford et al., 1998(6)	Mechanical LPB less than 6 weeks duration; referred by GP to physiotherapy; n=226;	Over time	Individual	Important: Average of clinician and patient global rating of important change (>5 on 15 point scale	Mean: 5 ROC area: 0.84 Initial scores: 0-8, 5-12, 9-16, 13-20, 17-24 estimates of important change: 2, 4, 5, 8, 8, respectively.	3 to 6 weeks
Kopec et al.,1995(21)	LBP patients seeking care from various treatment centers (physio, physiatry, orthopedic, rheumatology, pain clinics); n=242;	Over time	Group	Estimated (Time 2: 2 to 4 days): (self-rated: back pain "better, worse, or about the same") (Time 3: 6 months): (any positive change in ADL on 15 point scale)	Time 1 to 2: Effect size = 0.59; N&S sensitivity coeff = 0.20; Time 1 to 3: Effect size = 1.01; Correlation = 0.47; N&S sensitivity coeff = 0.35	Time1: baseline Time 2: 2 to 4 days (within 2 weeks), Time 3: 6 months (within 80 days)
Riddle et al., 1998(7)	Mechanical LPB less than 6 weeks duration; referred by GP to physiotherapy; n=143;	Over time	Individual	Important: Achievement of treatment goals	Mean: 5 ROC area: 0.68 Initial scores: 0-8, 5-12, 9-16, 13-20, 17-24 estimates of important change:3, 5, 8, 11, 13, respectively.	Initial visit to discharge

Appendix 2. Oswestry Disability Index.

Author	Patient Population	Which comparison	Who is focus	Type of change	Estimate of change (%)	Interval
Morrison et al., 1988(18)	LBP patients admitted to community hospital treatment program; n=120	Over time	Group	Observed	Mean change 6.0	Post-treatment, 1 year
Taylor et al., 1999(22)	LBP and sciatica patients; recruited from orthopedic clinic in tertiary care hospital; n=318	Over time	Group	Observed: 1. Treatment effect Estimated (patient's perceived global rating): 2. Better 3. Worse 4. No change	Median change: 1. 8.0 Mean change: 2. 18.4 3. -8.6 4. 2.1 Effect size (Cohen): 2. 1.1 3. 0.4	physio/ injection - 2 months; surgery - 6, 12, 24 months
Beurskens et al., 1996(9)	Non-specific LBP >6 weeks; n=81; treatment = traction.	Over time	A. Group B. Individual	Estimated: Patient global perceived effect	A. Mean: 11.9 Effect size: 0.80 B. ROC area: 0.76 (change 4-6 points)	5 weeks

Author	Patient Population	Which comparison	Who is focus	Type of change	Estimate of change (%)	Interval
Malmivaara et al., 1995(16)	Acute non-specific LBP; employees; n=186; treatment=bed rest, exercises, control	Between group, within person	Group	Observed difference: a. bed rest b. exercises	Mean (95% CI): 3 weeks a. 3.9 (-0.2, 8.0) b. 6.6 (2.0, 11.1) 12 weeks a. 3.8 (0.1, 7.5) b. 2.6 (-1.6, 6.7)	3, 12 weeks
Leclaire et al., 1996(15)	Work related LBP <3 months; Private physiatric outpatient clinic attendants; n=168; physiotherapy treatment or physio + back school	Over time	Groups	Observed: Treatment effect	Difference between groups (95% CI): Post-treatment 4.49 (.87, 8.12) No significant differences at 6, 12 months	Post-treatment, 6, 12 months
Stratford et al., 1994(10)	Mechanical LPB; referred by GP to physiotherapy; n=88;	Over time	A. Group B. Individual	Observed: 1. Treatment effect Important: 2. Average of clinician and patient global rating, importance of change (>5 on 15 point scale)	Mean (sd): 1. 16.1 (15.9) Correlation with criterion change: 2. 0.57 ROC analysis: 2. ROC area=0.78	4 to 6 weeks
Kopec et al., 1995(21)	LBP patients seeking care from various treatment centers (physio, physiatry, orthopedic, rheumatology, pain clinics); n=242;	Over time	Group	Estimated (Time 2: 2 to 4 days): (self-rated: back pain "better, worse, or about the same") (Time 3: 6 months): (any positive change in ADL on 15 point scale)	Time 1 to 2: Effect size = 0.50; N&S sensitivity coeff = 0.15; Time 1 to 3: Effect size = 0.61; Correlation = 0.35; N&S sensitivity coeff = 0.07	Time1: baseline Time 2: 2 to 4 days (within 2 weeks), Time 3: 6 months (within 80 days)

REFERENCES

1. Deyo RA, Battie M, Beurskens AJ, et al. Outcome measures for low back pain research. A proposal for standardized use. Spine 1998;23:2003-13.

2a. Roland M, Fairbank JC. Spine 2000; (in press).

2b. Roland M, Morris R. A study of the natural history of back pain. I. Development of a reliable and sensitive measure of disability in low-back pain. Spine 1983;8:141-4.

3. Fairbank JC, Couper J, Davies JB, O'Brien JP. The Oswestry Low Back Pain Disability Questionnaire. Physiotherapy 1980;66:271-3.

4a. Beaton DE, Bombardier C, Katz JN, et al. Looking for important change/differences in studies of responsiveness. J Rheumatol 2001;28:400-5.

4b. Stratford PW, Binkley J, Soloman P, Finch E, Gill C, Moreland J. Defining the minimum level of detectable change for the Roland-Morris questionnaire. Phys Ther 1996;76:359-68.

5. Dionne CE, Von Korff M, Koepsell TD, Deyo RA, Barlow WE, Checkoway H. A comparison of pain, functional limitations, and work status indices as outcome measures in back pain research. Spine 1999;24:2339-45.

6. Stratford PW, Binkley JM, Riddle DL, Guyatt GH. Sensitivity to change of the Roland-Morris back pain questionnaire. Part 1. Phys Ther 1998;78:1186-96.

7. Riddle DL, Stratford PW, Binkley JM. Sensitivity to change of the Roland-Morris back pain questionnaire: Part 2. Phys Ther 1998;78:1197-207.

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

9. Beurskens AJHM, deVet HCW, Koke AJA. Responsiveness of functional status in low back pain: A comparison of different instruments. Pain 1996;65:71-6.

10. Stratford PW, Binkley J, Solomon P, Gill C, Finch E. Assessing change over time in patients with low back pain. Phys Ther 1994;74:528-33.

11. Burton AK, Waddell G, Tillotson KM, Summerton N. Information and advice to patients with back pain can have a positive effect: a randomized controlled trial of a novel educational booklet in primary care. Spine 1999;24:2484-91.

12. Cherkin DC, Deyo RA, Battie M, Street J, Barlow W. A comparison

of physical therapy, chiropractic manipulation, and provision of an educational booklet for the treatment of patients with low back pain. N Eng J Med 1998;339:1021-9.
13. Constant F, Collin JF, Guillemin F, Boulange M. Effectiveness of spa therapy in chronic low back pain: a randomized clinical trial. J Rheumatol 1995;22:1315-20.
14. Constant F, Guillemin F, Collin JF, Boulange M. Use of spa therapy to improve the quality of life of chronic low back pain patients. Med Care 1998;36:1309-14.
15. Leclaire R, Esdaile JM, Suissa S, Rossignol M, Proulx R, Dupuis M. Back school in a first episode of compensated acute low back pain: a clinical trial to assess efficacy and prevent relapse. Arch Phys Med Rehabil 1996;77:673-9.
16. Malmivaara A, Hakkinen U, Aro T, et al. The treatment of acute low back pain — bed rest, exercises, or ordinary activity? N Eng J Med 1995;332:351-5.
17. Lanier DC, Stockton P. Clinical predictors of outcome of acute episodes of low back pain. J Fam Pract 1988;27:483-9.
18. Morrison GE, Chase W, Young V, Roberts WL. Back pain: treatment and prevention in a community hospital. Arch Phys Med Rehabil 1988;69:605-9.
19. Patrick DL, Deyo RA, Atlas SJ, Singer DE, Chapin A, Keller RB. Assessing health-related quality of life in patients with sciatica. Spine 1995;20:1899-909.
20. Sinclair SJ, Hogg-Johnson SA, Mondloch MV, Shields SA. The effectiveness of an early active intervention program for workers with soft tissue injuries: the Early Claimant Cohort study. Spine 1997;22:2919-31.
21. Kopec JA, Esdaile JM, Abrahamowicz M, et al. The Quebec Back Pain Disability Scale measurement properties. Spine 1995; 20:341-52.
22. Taylor SJ, Taylor AE, Foy MA, Fogg AJB. Responsiveness of common outcome measures for patients with low back pain. Spine 1999;24:1805-12.