# Pain in low-back pain

## Problems in measuring outcomes in musculoskeletal disorders

**Gustavo Zanoli[1,2], Björn Strömqvist[2], Bo Jönsson[3], Roberto Padua[4] and Emilio Romanini[5]**

Departments of Orthopedics, [1] University of Ferrara, Ferrara, Italy, [2] Lund University Hospital, Lund, Sweden, [3] Helsingborg, Sweden, [4] Ospedale S. Giacomo, Rome, Italy, and [5] ISPEO, Rome, Italy.
Correspondence: Gustavo Zanoli, Department of Orthopaedics, University of Ferrara, Arcispedale S. Anna, 44100 Ferrara, Italy.
Tel +39 0532 236573. Fax +39 0532 209250. E-mail gzanoli@globeweb.org

One of the crucial issues of clinical research in musculo-skeletal disorders is outcome measurement. This has been the central question of the Outcomes Movement, and has occupied many researchers in all musculo-skeletal specialties: which outcomes should we measure? Rheumatologists, Rehabilitation Physicians, Traumatologists and Orthopeadic Surgeons have tried to answer this question from different points of view, very often isolated from each other and without interpolating efforts.

One good example of a multidisciplinary subject is low-back pain (LBP). It is not a disease, but a symptom, and a common experience in the life of almost every human being, as well as a growing cause of direct and indirect costs for the social systems in many industrialized countries. From a medical point of view, it is a multifaceted syndrome: causes and clinical presentations vary widely. Psychological, social and economic status of the patient seem to play a role in the history of LBP, and many different specialists and disciplines are involved in the treatment.

To compare results, we would need a simple and standardised outcome measure that accounts for different clinical presentations, interactions, and therapeutic approaches (Deyo et al. 1998). Unfortunately, as in most musculoskeletal disorders, very rarely simple dichotomous outcomes can be used in evaluating results of treatment alternatives for LBP. Patients rarely are totally ill or completely disease-free, but very often fluctuate between better or worse health status.

Since we are talking about Low-Back Pain, pain is probably the most straightforward outcome to measure in the evaluation of therapeutic interventions for this condition. We decided to use it as an example in considering critical aspects of the choice, the use, and the interpretation of an outcome measure in musculo-skeletal clinical research.

### Problems of choice

Evidence-based medicine, which is the guiding line of this supplement, deals with the use of the best available evidence in providing healthcare to patients. But sometimes we might want to know evidence from our own practice (outcomes assessment), or even to produce evidence that can be useful to others and make it available through publications. In both cases we need to choose a measure by which we quantify the health status of persons at different stages of the clinical evolution. In a recently published Spine Focus Issue, dedicated to Outcome Assessments in the Evaluation of Treatment of Spinal Disorders, the Key Questions to Ask About an Outcome Measure were summarised (Bombardier 2000). Researchers in the field of LBP might find a lot of useful information in this volume. In the same issue, a systematic review of the literature in the search for a Health-Related Quality of Life (HRQoL) instrument to be used in LBP patients was presented (Zanoli et al. 2000), the idea behind which

was to apply an evidence-based approach in the choice of an outcome measure. More than 90 different instruments used during the past 30 years in approximately 500 published clinical papers were identified. If we restrict the scope to those who measure specifically pain, about 30 instruments remain. These are still too many to imagine the possibility of determining the best one for the researcher with an evidence-based methodology. Most likely there not one single best method of assessing pain, but different instruments are appropriate in different settings or study designs. However, in the perspective of an evidence-based practitioner, these 30 measures can be useful to raise some methodological issues to consider when reading scientific reports on LBP that include a measure of pain.

## Methodological considerations

We have limited the topic of this presentation to patient-oriented measures, not only because other types of so-called "objective" measures of pain, such as laboratory tests, usually lack a sound scientific background as well as clinical utility, but also because self-reported questionnaires are becoming the standard method of evaluation in musculoskeletal care, and many of the following considerations can be applied to similar outcome measures in fields other than LBP.

The most straight forward to measure pain is probably simply asking the patient to quantify it on visual analog scale (VAS) (Huskisson 1974) or on a numerical radical scale (NRS) (Jensen et al. 1986). Both approaches have been used more than once in LBP (Evans and Kagan 1986, Schofferman 1999). Another even more commonly used approach is to submit to the patient an ordinal scale of attributes that describe pain (for example no pain, mild pain, some pain, a lot of pain, worst pain), and goes under the name of Verbal Rating Scale (VRS): we should probably speak of verbal rating scales, as selection of words and number of possible choices in the ordinal scale vary from study to study. In a methodological perspective these scales can be considered as patient-oriented measures, even though it is likely that before the dissemination of knowledge about self-administered questionnaires, VAS, NRS and VRS were actually collected by physicians in many studies.

From a formal point of view, these are single items, as opposed to the more complex multi-items questionnaires, which have been developed more recently. Out of 15 specific pain questionnaires found in our review, 5 were personal set of question, designed for a specific study, 5 were proposed questionnaires, which did not undergo further evaluation, and 5 were validated questionnaires, which underwent psychometric or clinimetric testing also in LBP settings. Reliability (the ability of a given instrument to give reproducible measures of the measured entity), validity (the ability of a given instrument to give "true" measures of the measured entity), and responsiveness (the ability of a given instrument to detect a known change in the measured entity), are usually studied in these undertakings, which very often provide also useful normative data to enable correlation with different population samples and power calculation. For these reasons, and to encourage the use of very few "standards" which allow comparisons across studies, validated questionnaire are preferable.

If we concentrate on the substance of measurement, we can assume that, unless otherwise specified, VAS, and NRS are intended to be measures of pain intensity. It is important, when trying to extract evidence from an article, to check if this was actually the case. For the same reason it would be important that authors specify what they mean by VAS measurement. VRS usually clarify this aspect with the choice of the terms included in the administered scale. Measuring pain intensity seems to be by far the most common approach to the quantification of pain, but is not the only one. Pain-related disability can be connected with pain intensity to assess pain severity. This happens for example in the SF-36 bodily pain domain, which combines a question on pain intensity with one on interference with activities.

More recently patient-specific questionnaires have been introduced, which include self-rated importance of symptoms or items to calibrate the results (Chatman et al. 1997). A different approach to quantification of pain, can be the assessment of number of days (hours) with pain. This can be done either retrospectively or in a prospective manner (diary). These methods can be very important in

the case of chronic and/or recurrent pain, two situations which are quite different from the acute perspective, and for which we still lack accepted definitions and dedicated instruments. In fact, most of the measures in use have been subsequently adapted from the acute setting where they were originally validated, and it has been shown that they posses different measurement properties in the various situations (see below).

Other aspects of the experience of pain can be taken into account, such pain affect, which may be defined as the ensemble of emotions and feelings that accompany the sensation of pain itself. For a more in-depth review of the evidence on this subject, see (Von Korff et al. 2000).

Pain can be determined absolutely, i.e. prospectively recording the measure at different time points and then calculating the difference, or relatively, asking the patient to judge how much his symptoms have decreased or increased after a defined amount of time. The latter solution does not imply that the study is retrospective, since the choice of this outcome measure could have been prospectively included in the study design, but leaves us without any data to characterise patient population at the beginning of the study. Among others, we have recently shown that also in LBP these two approaches do not necessarily lead to the same result (Zanoli et al. 2001). Different types of recall bias have been often advocated to explain differences between the two methods, suggesting that the absolute prospective one is the correct one (Linton and Melin 1982, Ross 1989, Streiner and Norman 1989). Even though widely supported, differential scores are less immediate as outcome measures because they introduce arbitrary calculation, which might increase sources of variability and error. Besides, asking the patients if he is feeling better is probably the most ancient of all outcome measures, and a positive answer is the aim of all our efforts in everyday practice. Most likely, in our opinion, absolute and relative measures represent different aspects of the picture, and should be evaluated separately.

## Interpretation

These considerations introduce the more philosophical aspects of interpretation of pain measures. This is a very controversial issue, but a crucial one, especially when we try to practice evidence-based musculoskeletal care. First of all, as we have seen, very often apparently similar methods of assessing pain cannot be assimilated, even less pooled. Not all systematic reviews address this issue properly (Furlan et al. 2001). Cochrane back group has given attention to outcome evaluation in their methodological guideline for reviewer (Van Tulder et al. 1997). In the field of pain, possible solutions have been proposed (Moore et al. 1997). In our opinion, an arbitrary data extraction and pooling of results very often is due to a misinterpretation of the differences between outcome evaluation across trials, and usually constitute the only major flaw in systematic reviews, especially when authors lack direct clinical experience in the field.

Second, we must consider the difference between statistical significance and clinical significance. Large samples most likely produce statistically significant results, but the entity of the result can be irrelevant from the patient's perspective. For this reason, Minimally Clinically Important Differences (MCIDs) have been introduced and defined in various settings as well as in LBP (Beaton 2000). These can be useful not only to weigh the results, but also for sample size calculations in the planning phase of a new study. It is important to remember that different clinical settings may imply different MCIDs: for example, MCID for VAS for acute pain (9-13 mm) could be different from that of chronic pain (20 mm), and MCID for patients with little pain at enrolment might differ from that of most suffering patients (Todd et al. 1996, Kelly 1999, Bird and Dickson 2001)

A final difficulty occurs when we need to express the clinical value of the studies to our patients, for example to quantify our own or our patient's NNTs or NNHs. After having checked that the paper we analyse has used a sound methodology in outcome evaluation, we must also be certain of the implication of the outcome considered, to decide if it applies to our patients, and include the result of that study in our shared-decision-making process.

In conclusion, clinical researchers should bear in mind that it is necessary to be able to refer to few, valid and recognised standard methods of outcome evaluation, in order to speed up this process

and allow quick understanding and exchangeable comparisons across studies. In the end, the role and responsibility of every single practicing physician is crucial in this process: translating exoteric acronym such as VAS or MCIDs into useful patient information is one the things that makes medicine still an art in the evidence-based era.

## References


Beaton DE. Understanding the relevance of measured change through studies of responsiveness. Spine 2000; 25(24): 3192-9

Bird SB, Dickson EW. Clinically significant changes in pain along the visual analog scale. Ann Emerg Med 2001; 38(6): 639-43.

Bombardier C. Outcome assessments in the evaluation of treatment of spinal disorders. Introduction. Spine 2000; 25(24): 3097-9.

Chatman AB, Hyams SP, Neel JM, Binkley JM, Stratford PW, Schomberg A, Stabler M. The Patient-Specific Functional Scale: measurement properties in patients with knee dysfunction. Phys Ther 1997; 77(8): 820-9

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(18): 2003-13.

Evans JH, Kagan A 2d. The development of a functional rating scale to measure the treatment outcome of chronic spinal patients. Spine 1986; 11(3): 277-81.

Furlan AD, Clarke J, Esmail R, Sinclair S, Irvin E, Bombardier C. A critical review of reviews on the treatment of chronic low back pain. Spine 2001; 26(7): E155-62

Huskisson EC. Measurement of pain. Lancet. 1974; 2(7889): 1127-31.

Jensen MP, Karoly P, Braver S. The measurement of clinical pain intensity: a comparison of six methods. Pain 1986; 27: 117–26

Kelly AM. Does the clinically significant difference in visual analog scale pain scores vary with gender, age, or cause of pain? Acad Emerg Med 1998; 5(11): 1086-90.

Linton SJ, Melin L The accuracy of remembering chronic pain. Pain 1982; 13(3): 281-5

Moore A, Moore O, McQuay H, Gavaghan D. Deriving dichotomous outcome measures from continuous data in randomised controlled trials of analgesics: use of pain intensity and visual analog scales. Pain 1997; 69(3): 311-5

Ross M Relation of implicit theories to the construction of personal histories. Psych Rev 1989; 96:341-57

Schofferman J. Long-term opioid analgesic therapy for severe refractory lumbar spine pain. Clin J Pain 1999; 15(2): 136-40

Streiner DL, Norman GR. Health measurement scales. A practical guide to their development and use. 2nd Ed. New York: Oxford University Press; 1989.

Strömqvist B, Jönsson B. Computerized follow-up after surgery for degenerative lumbar spine diseases. Acta Orthop Scand 1993; Suppl 251: 138-42

Todd KH, Funk KG, Funk JP, Bonacci R. Clinical significance of reported changes in pain severity. Ann Emerg Med 1996; 27(4): 485-9.

van Tulder MW, Assendelft WJ, Koes BW, Bouter LM. Method guidelines for systematic reviews in the Cochrane Collaboration Back Review Group for Spinal Disorders. Spine 1997; 22(20): 2323-30

Von Korff M, Jensen MP, Karoly P. Assessing global pain severity by self-report in clinical and health services research. Spine 2000; 25(24): 3140-51.

Zanoli G, Strömqvist B, Jönsson B. Visual analog scales for interpretation of back and leg pain intensity in patients operated for degenerative lumbar spine disorders. Spine 2001; 26(21): 2375-80

Zanoli G, Strömqvist B, Padua R, Romanini E. Lessons learned searching for a HRQoL instrument to assess the results of treatment in persons with lumbar disorders. Spine 2000; 25(24): 3178-85