SPINE Volume 35, Number 21, pp E1115–E1119

The Visual Analog Scale and a Five-Item Verbal Rating Scale Are Not Interchangeable for Back Pain Assessment in Lumbar Spine Disorders

Antònia Matamalas, MD,* Manuel Ramírez, MD,* Sergi Mojal, Licentiate in Biostatistics,† Ana García De Frutos, MD,* Antonio Molina, MD, PhD,* Guillem Saló, MD, PhD,* Andreu Lladó, MD, PhD,* and Enric Cáceres, MD, PhD*

Study Design. Prospective study of patients with chronic back pain from lumbar spine disorders.

Objective. To evaluate the degree of interchangeability of a 100-mm visual analog scale (VAS) and a 5-point verbal rating scale (VRS) for the assessment of pain intensity.

Summary of Background Data. The fact that VAS and Likert scales are highly intercorrelated does not mean that both types of scales can be used interchangeably.

Methods. A total of 151 patients (mean age, 52 ± 14.6 years) undergoing elective spine surgery completed a 100-mm VAS and a discrete 5-category VRS corresponding to the first item question of the core set ("How severe was your back pain in the last week?"). Pain intensity on the VAS was rated using the same question than for the VRS. The level of order-consistency (monotonic agreement), disordered pairs (D), percentage of agreement, and systematic disagreement (relative position), and relative concentration ([RC]) were estimated. VAS assessments were transformed into a discrete 5-category, with the cut-off VAS positions being defined by quintiles and equidistantly.

Results. For VAS defined equidistantly, monotonic agreement was 0.840, D was 0.080, and the percentage of identical pairs was 53%. The corresponding figures for VAS defined by quintiles were 0.809, 0.096, and 27.8%. Inconsistencies between the VAS and the VRS scales were also demonstrated by the marginal distributions, with PR values of −0.005 (95% confidence interval [CI], −0.011 to −0.002) and RC values of 0.144 (95% CI, 0.137–0.152) for VAS defined equidistantly, and PR values of 0.391 (95% CI, 0.384–0.397) and RC values of 0.265 (95% CI, 0.255–0.275) for VAS defined by quintiles.

Conclusion. The order-consistency level was low with overlapping of pain records between the 2 scales, indicating that VAS and VRS are not interchangeable and, therefore, a results obtained with the use of each scale cannot be compared.

Key words: low back pain, visual analog scale, 5-point verbal rating scale, pain intensity, lumbar degenerative disease. **Spine 2010;35:E1115–E1119**

Reliable and accurate assessments of perceived low back pain are indispensable in clinical practice for diagnosis, selection of the most appropriate treatment modality, and evaluation of treatment efficacy in patients with degenerative disc disease.[1] Pain is a multifactorial unpleasant sensory experience, which is influenced by different aspects, such as the intensity of pain and the affective component.[2] Pain intensity or pain severity is probably the most frequently assessed dimension.[3,4] The most commonly used scales for measuring pain intensity are the visual analog scale (VAS) and the verbal rating scales (VRS). VAS consists of a 100-mm straight horizontal (or vertical) line labeled "no pain" at one end and "pain" as bad as could possible be on the other.[5] In contrast to the continuous VAS, Likert-type scales are comprised of a variable number of word categories (*e.g.*, 2-point VRS and 11-point numerical rating scale), although the number of levels needed to assess self-reported pain intensity is an important issue that has yet to be resolved.[6,7]

Both VAS and Likert scales have been evaluated in terms of reliability, validity, and responsiveness. In general, both scaling methods seem to be reliable and valid for its use in routine clinical assessment.[8–12] It has been shown that the correlation between both scales is highly significant and most authors consider that both types of scales measure the same aspects of pain and can be used interchangeably. Therefore, VAS and Likert scales are indistinctively used for direct comparison of the results obtained in different studies. However, the fact that 2 scales are highly intercorrelated does not necessarily mean that they measure the same construct. Moreover, considerable interindividual variability and overlap of VAS scores and VRS ratings have been reported, and there is no enough data to conclude that one scale type is better than the other.[1,12,13]

In spine patients, the scarce evidence on the results of different treatment methods of degenerative lumbar spine diseases requires the use of standardized outcome measures to allow international comparisons.[14] In this

From the *Service of Orthopaedic Surgery and Traumatology, Hospital Universitario del Mar, Universitat Autònoma de Barcelona, Barcelona, Spain; and †Service of Methodological Consulting in Biomedical Research, Institut Municipal d'Investigació Mèdica (IMIM-Hospital del Mar), Barcelona, Spain.

Acknowledgment date: February 24, 2009. First revision date: June 26, 2009. Second revision date: March 8, 2010. Acceptance date: March 12, 2010.

Presented as an oral communication at: The 9th Congress of the European Federation of National Associations of Orthopaedics and Traumatology (EFORT), Nice, France. 29 May–1 June, 2008.

The manuscript submitted does not contain information about medical device(s)/drug(s).

No funds were received in support of this work. No benefits in any form have been or will be received from a commercial party related directly or indirectly to the subject of this manuscript.

Address correspondence and reprint requests to Manuel Ramírez, MD, Service of Orthopaedic Surgery and Traumatology, Hospital Universitario del Mar, Universitat Autònoma de Barcelona, Passeig Marítim 25–29, E-08003 Barcelona, Spain; E-mail: mramirez@imas.imim.es

respect, the disparity in results published on the interchangeability of VAS and VRS scales and the need for standardized measurement tools make necessary to determine to what extent both scales can be compared for the assessment of results in lumbar degenerative disease. The objective of this study was to evaluate the degree of interchangeability of VAS and a 5-point Likert scale for back pain assessment in patients with lumbar spine disorders undergoing elective spine surgery.

■ Materials and Methods

Between June 2007 and July 2008, all consecutive patients with lumbar degenerative disease scheduled to undergo elective spine surgery at the Spinal Unit of the Department of Orthopedic Surgery and Traumatology of Hospital Universitario del Mar in Barcelona, Spain, were eligible to participate in the study. Illiterate patients and those who were unable to self-complete the study questionnaires were excluded from the study as were those who refused to participate. The assessments were conducted in accordance with the declaration of Helsinki and written informed consent was obtained from all participants. The study was approved by the Ethics Committee of Hospital Universitariodel Mar.

All patients were admitted to the hospital 24 hours before surgery, and were administered a series of patient-based outcome assessments instruments, including the Spanish version of the Oswestry Low Back Pain Disability Questionnaire,[15,16] the pain item of the core set developed by Deyo *et al*,[17,18] the SF-36 Health Survey,[19] a VAS for pain intensity, and a questionnaire on work status. The discrete 5-category VRS corresponded to the first item question of the core set (How severe was your back pain in the last week?), with the eligible alternatives of "none," "mild," "moderate," "severe," and "very severe." The VAS scale was a 100-mm long horizontal line labeled no pain at one end and "worst pain possible" at the other end. In order to avoid bias in the wording for the patients' experienced pain intensity level, they were asked to rate their pain intensity on the VAS using the same question than for the VRS.

Statistical Analysis

In order to describe the correspondence between VAS data and the VRS categories, the continuous VAS assessments were transformed into a discrete 5-category scale in 2 ways, the cut-off positions on the VAS being defined by quintiles (when dividing the VAS scores into 5 groups according to the percentage of patients for each score) and equidistantly (by condensing the continuous VAS records into an equidistant 5-category scale that is to be compared with the 5-category VRS). Given that each individual assessed their perceived pain on 2 scales, the dataset consisted of paired data.[1] The analysis of data was carried out according to the statistical method described by Svensson and Holm.[20] Interchangeability between scales with different numbers of response categories requires a high level of order-consistency, *i.e.*, lack of overlapping of the records of 1 scale relative the other. According to the method of Svensson,[20,21] the level of order-consistency between scales with the same number of categories requires a high percentage of agreement, *i.e.*, a high proportion of identical pairs, and a lack of systematic disagreement (bias) by the absence of different frequency distributions, also called marginal distributions. Pairs were considered overlapping when the responses obtained on the 2 scales were found in same level of pain, and not overlapping when responses obtained on the 2 scales were found in different levels of pain. In addition, pairs should be ordered, that is, low scores on the VAS should be consistent with low scores on the VRS. The number of disordered pairs was calculated and defined the measure of disorder, D. Systematic disagreement was measured by means of the relative position (RP) and the relative concentration (RC) with possible values ranging from −1 to 1. The RP estimates the probability of the pain assessments on 1 scale being shifted toward higher or lower categories relative to the other. The RC estimates the differences between the probability of the pain assessments on 1 scale being concentrated to the other and *vice versa*. The distribution of the pairs of data was evaluated by means of 5×5 contingency tables[13] (Table 1). The level of order-consistency was estimated by the coefficient of monotonic agreement (MA), which is defined by the difference between the probabilities of ordered and disordered pairs of assessments. The range of possible values of MA is ×1 to 1. The degree of concordance between VAS and VRS instruments was also assessed with kappa statistics. The R statistical package was used for data analysis.

Table 1. Frequency Distributions of Paired Data of Condensed 5-Category VAS *Versus* the Discrete VRS in Case of Maximum Order-Consistency

	VAS					
VRS	1	2	3	4	5	
None	1	1	0	0	0	2
Mild	0	5	2	2	2	10
Moderate	1	2	9	8	1	21
Severe	0	1	8	33	20	62
Very severe	0	1	2	17	36	56
	2	10	21	60	58	151

VAS indicates visual analog scale; VRS, verbal rating scale.

■ Results

There were 293 patients undergoing spine surgery during the study period. Of these patients, 163 underwent lumbar spine surgery but 12 were excluded (8 refused to take part in the study and 4 immigrant patients had difficulties in understanding the Spanish language). The final study population included 151 patients who agreed to participate in the study and completed both the VAS and the VRS instruments. There were 72 men and 79 women, with a mean (standard deviation) age of 52.2 (14.6) years. The etiology of back pain was degenerative disc disease (ICD-9, code 722.52: degenerative disc disease) in 43 patients; lumbar canal stenosis (ICD-9, code 724.02: spinal stenosis, lumbar region) in 32; pseudarthrosis (ICD-9, code 77.18: reinterventions after other procedures) in 26; disc herniation (ICD-9, code 722.1: lumbar disc displacement) in 22; spondylolisthesis(ICD-9, code: 756.12: spondylolisthesis) in 17; and other conditions (ICD-9, code 724: other and unspecified disorders of back) in 11.

The measured level of concordance was MA = 0.840 for VAS defined equidistantly and MA = 0.809 for VAS defined by quintiles, revealing a difference between the

Table 2. Frequency Distributions of Paired Data of Condensed 5-Category VAS Defined Equidistantly *Versus* the Discrete VRS

	VAS					
VRS	1	2	3	4	5	
None	1	4	0	1	0	6
Mild	1	3	5	2	2	13
Moderate	0	1	10	13	0	24
Severe	0	1	3	23	13	40
Very severe	0	1	3	21	43	68
	6	13	21	60	58	151

Identical Pairs (n = 80) are Distributed From the Upper-Left to Lower-Right in the Main Diagonal.
VAS indicates visual analog scale; VRS, verbal rating scale.

Table 4. Inter-Scale Comparisons of VAS *Versus* VRS for the Assessment of Back Pain Intensity in 151 Patients With Lumbar Degenerative Disease

Measure	Equidistant VAS *vs.* VRS	VAS Defined by Quintiles *vs.* VRS
MA	0.840	0.809
D (95% CI)	0.080 (0.054–0.107)	0.096 (0.064–0.127)
PA	0.530	0.278
RP (95% CI)	−0.005 (−0.011 to −0.002)	0.391 (0.384–0.397)
RC (95% CI)	0.144 (0.137–0.152)	0.265 (0.255–0.275)

Monotonic agreement (MA) indicates level of order-consistency.
Measure of disorder (D) indicates disordered pairs.
Percentage agreement (PA) indicates proportion of identical pairs.
Relative concentration (RC) indicates systematic disagreement.
VAS indicates visual analog scale; VRS, verbal rating scale; CI, confidence interval; RP, relative position.

ordered and disordered pairs of assessments. With regard to the measure of disorder, D, values of 0.080 (95% confidence interval [CI], 0.054–0.107) were obtained for VAS defined equidistantly and 0.096 (95% CI, 0.064–0.127) for VAS defined by quintiles. A low level of order-consistency between both scales was observed. The percentage of identical pairs was 53% (80/151) for VAS scores divided into a 5-equidistant category scale and 27.8% (42/151) for VAS defined by quintiles (Tables 2, 3). Results of kappa statistics for the equidistant and quintiles approaches were $\kappa = 0.0425$ (95% CI, 0–0.125) and $\kappa = 0.297$ (95% CI, 0.183–0.412), respectively.

Inconsistencies between the VAS and the VRS scales were also demonstrated by the marginal distributions, with PR values of −0.005 (95% CI, −0.011 to −0.002) and RC values of 0.144 (95% CI, 0.137–0.152) for VAS defined equidistantly, and PR values of 0.391 (95% CI, 0.384–0.397) and RC values of 0.265 (95% CI, 0.255–0.275) for VAS defined by quintiles. Interscale comparisons of VAS *versus* VRS are summarized in Table 4.

■ Discussion

Pain, in general, and back pain, in particular, constitute a subjective experience that can be influenced by multiple variables.[2,22,23] Back pain is often multifactorial, making assessment of pain intensity or the perceived severity of felt pain difficult. Although comprehensive assessment of pain includes a thorough evaluation of the patient's pain and its impact on physical and psychosocial functions, as well as other aspects of quality of life, evaluation of pain severity is an essential component that guides treatment planning. VAS and VRS are the most commonly used pain rating scales in daily practice.[13,24,25] Different studies have compared the reliability, validity, and/or sensitivity to changes in pain intensity of the VAS and the VRS scales.[8–11,24,26] Although it appears that sensitivity is influenced by conditions in which pain intensity is measured,[27–30] some authors consider that both types of scales are comparable in terms of reliability and validity, and provide similar results.[31] Therefore, although both scales are valid for the assessment of pain intensity, there are no sufficient data to allow the use of one scale type because it is better than the other.[32] However, VAS and VRS are frequently used indistinctively without definite certainty that both scales are interchangeable. In fact, in clinical practice, VAS scores are frequently systematically categorized as discrete variables and many physicians consider VAS between 0 and 3 as mild pain, between 3 and 6 as moderate pain, and >7 as severe pain.

Table 3. Frequency Distributions of Paired Data of Condensed 5-Category VAS Defined by Quintiles *Versus* the Discrete VRS

	VAS					
VRS	1	2	3	4	5	
None	2	8	10	8	2	30
Mild	0	1	7	13	4	25
Moderate	0	0	1	18	9	28
Severe	0	1	1	15	20	37
Very severe	0	0	2	6	23	31
	2	10	21	60	58	151

Identical Pairs (n = 42) are Distributed From the Upper-Left to Lower-Right in the Main Diagonal.
VAS indicates visual analog scale; VRS, verbal rating scale.

Different statistical approaches have been used for the comparison of VAS and VRS scales.[33–35] The Cohen coefficient of agreement[36] for nominal scales is the most widely used measure of agreement between categorical data; however, neither the order nor the magnitude of disagreement can be assessed by the kappa statistics. In contrast, the method described by Svensson[21] is more appropriate to assess the interchangeability between scales, because it shows a more exhaustive analysis of the level of disagreement based on the presence or absence of different marginal distributions, calculated by the RP and the RC. In order to consider 2 scales with different numbers of response categories as interchangeable, not only good concordance between scales is required, but also a high level of order-consistency.[1,21]

In the present study, the pain item of the core set developed by Deyo *et al*,[17,18] was used as the VRS. This questionnaire has been shown to be reliable, valid, and

easy to be administered in patients with subacute osteoporotic fractures and chronic back pain.[6,17,18] VAS was formulated using the same question than that in the pain item of the core set in order to avoid interpretation bias. The present findings indicate a strong correlation between both scales, with MA values of 0.804 and 0.809 for VAS defined equidistantly and by quintiles, respectively. In the study of Hasson and Arnetz,[12] moderate to strong correlations were found between single-item VAS and a single-item Likert question. However, when the order-consistency level was analyzed, the percentage of identical pairs was low either for VAS scores defined equidistantly or by quintiles. This means that there is a high overlapping of the records on 1 scale relative to the other, which is in accordance with the overlap of numerical scores of VAS when plotted against a 4-point simple descriptive scale found in the study of Downie *et al*.[37] Moreover, there was a systematic disagreement, particularly remarkable for VAS scores defined by quintiles, so that VAS scores were systematically higher than the scores of VRS, thereby, both scales cannot be regarded as interchangeable. These results are consistent with date reported by Svensson in previous studies.[13,21]

The present findings may be interpreted, in part, by the fact that it may be difficult for the patients to express the pain perceived as marks or numbers when they are unaware of their meaning, given that an operational definition of the levels of pain intensity is lacking.[13] In the case of VRS, it is also possible that the offered response options cannot correspond to the pain perceived by the patient, all of which will affect the response obtained. Overestimation of pain severity by VAS scores may be partly explained by the nonlinear characteristics of VAS,[1] so that a VAS score of 80 not necessarily means a 2-fold pain intensity compared with a VAS score of 40. Moreover, when VAS was categorized into in 5 equidistant categories, an important systematic inconsistency in the concentration of RC $\neq$ 0 was observed, determining a risk for overestimation or underestimation of the variables analyzed probably due to the impact of VAS nonlinearity on rescaling of the responses.[13] On the other hand, it has been shown that is unclear which point on the VAS scale is equivalent to pain of at least moderate intensity measured by a standard 4-point categorical scale.[38]

The present findings should be interpreted taking into consideration some limitations of the study, including the relatively small sample size and the fact that the study population was a selected group of patients with lumbar degenerative disease scheduled for elective spine surgery and do not represent frequent conditions in other settings, such as primary care and rehabilitation units. Therefore, generalization of the results obtained to patients with less severe back pain deserves further study. The fact that a homogeneous group of patients in terms of specific etiology of chronic back pain was evaluated is a strength of the study.

In conclusion, the VAS and the 5-point VRS were strongly correlated and both scales are valid to measure the intensity of pain in patients with chronic back pain from lumbar degenerative disease. However, the order-consistency level was low, with overlapping of pain records between the 2 scales, indicating that VAS and VRS are not interchangeable and, therefore, a comparison of the results obtained with the use of each scale is not methodologically adequate. Studies using VAS should be compared with studies using the same scale, whereas studies using VRS should be compared with those also using VRS.

■ Key Points

- Interchangeability of a 100-mm VAS and Likert-type VRS to assess pain intensity in patients with chronic back pain from lumbar spine disorders has not been established.
- Data obtained in a series of 151 patients who completed both scales using the first item (pain) question of the core set were analyzed to determine the level of order-consistency disordered pairs, percentage of identical pairs (agreement), and systematic disagreement between both scales.
- Although a strong correlation between both scales was found, the percentage of identical pair was low. Moreover, a low order-consistency level with overlapping of pain records between the 2 scales was observed, indicating that systematic disagreement was present.
- VAS and VRS are not interchangeable and, therefore, the results obtained independently with the use of each scale cannot be compared.

Acknowledgments

The authors thank J. M. Manresa and S. Mojal for statistical analysis, and Marta Pulido, MD, for editing the manuscript and editorial assistance.

References

1. Lund I, Lundeberg T, Sandberg L, et al. Lack of interchangeability between visual analogue and verbal rating pain scales: a cross sectional description of pain etiology groups. *BMC Med Res Methodol* 2005;5:31. doi: 10.1186/1471–2288–5–31.
2. Gatchel RJ, Polatin PB, Mayer TG. The dominant role of psychosocial risk factors in the development of chronic low back pain disability. *Spine* 2005;20:2702–9.
3. Mannion AF, Balagué F, Pellisé F, et al. Pain measurement in patients with low back pain. *Nat Clin Pract Rheumatol* 2007;3:610–8.
4. Breivik H, Borchgrevink PC, Allen SM, et al. Assessment of pain. *Br J Anaesth* 2008;101:17–24.
5. Wewers ME, Lowe NK. A critical review of visual analogue scales in the measurement of clinical phenomena. *Res Nurs Health* 1990;13:227–36.
6. Ohnhaus EE, Adler R. Methodological problems in the measurement of pain: a comparison between the verbal rating scale and the visual analogue scale. *Pain* 1975;1:379–84.
7. Jensen MP, Turner JA, Romano JM. What is the maximum number of levels needed in pain intensity measurement? *Pain* 1994;58:387–92.
8. Lara-Muñoz C, De Leon SP, Feinstein AR, et al. Comparison of three rating

scales for measuring subjective phenomena in clinical research. I. Use of experimentally controlled auditory stimuli. *Arch Med Res* 2004;35:43–8.
9. Ponce de Leon S, Lara-Muñoz C, Feinstein AR, et al. A comparison of three rating scales for measuring subjective phenomena in clinical research. II. Use of experimentally controlled visual stimuli. *Arch Med Res* 2004;35:157–62.
10. Baños JE, Bosch F, Cañellas M, et al. Acceptability of visual analogue scales in the clinical setting: a comparison with verbal rating scales in postoperative pain. *Methods Find Exp Clin Pharmacol* 1989;11:123–7.
11. Briggs M, Closs JS. A descriptive study of the use of visual analogue scales and verbal rating scales for the assessment of postoperative pain in orthopedic patients. *J Pain Symptom Manage* 1999;18:438–46.
12. Hasson D, Arnetz BB. Validation and findings comparing VAS vs. Likert scaler for psychosocial measurements. *Int Electron J Health Educ* 2005;8:178–92.
13. Svensson E. Concordance between ratings using different scales for the same variable. *Stat Med* 2000;19:3483–96.
14. Zanoli G. Outcome assessment in lumbar spine surgery. *Acta Orthop Suppl* 2005;76:5–47.
15. Fairbank JC, Couper J, Davies JB. The Oswestry low back pain questionnaire. *Physiotherapy* 1980;66:271–3.
16. Florez García MT, García Pérez MA, García Pérez F, et al. Adaptación transcultural a la población española de la escala de incapacidad por dolor lumbar de Oswestry [in Spanish]. *Rehabilitación* 1995;29:138–45.
17. 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.
18. Ferrer M, Pellisé F, Escudero O, et al. Validation of a minimum outcome core set in the evaluation of patients with back pain. *Spine* 2006;31:1372–9.
19. Alonso J, Regidor E, Barrio G, et al. Population reference values of the Spanish version of the health questionnaire SF-36 [in Spanish]. *Med Clin (Barc)* 1998;111:410–6.
20. Svensson E, Holm S. Separation of systematic random differences in ordinal rating scales. *Stat Med* 1994;13:2437–53.
21. Svensson E. Comparison of the quality assessments using continuous and discrete ordinal rating scales. *Biom J* 2000;42:417–34.
22. Hasenbring M, Marienfeld G, Kuhlendahl D, et al. Risk factors of chronicity in lumbar disc patients. A prospective investigation of biologic, psychologic, and social predictors of therapy outcome. *Spine* 1994;19:2759–65.
23. Pincus T, Vlaeyen JW, Kendall NA, et al. Cognitive-behavioral therapy and psychosocial factors in low back pain: directions for the future. *Spine* 2002;27:E133–8.
24. Herr KA, Spratt K, Mobily PR, et al. Pain intensity assessment in older adults: use of experimental pain to compare psychometric properties and usability of selected pain scales with younger adults. *Clin J Pain* 2004;20:207–19.
25. Williamson A, Hoggart B. Pain: a review of three commonly used pain rating scales. *J Clin Nurs* 2005;14:798–804.
26. Peters ML, Patijn J, Lamé I. Pain assessment in younger and older pain patients: psychometric properties and patient preference of five commonly used measures of pain intensity. *Pain Med* 2007;8:601–10.
27. Jensen MP, Miller L, Fisher LD. Assessment of pain during medical procedures: a comparison of three scales. *Clin J Pain* 1998;14:343–9.
28. Bolton JE, Wilkinson RC. Responsiveness of pain scales: a comparison of three pain intensity measures in chiropractic patients. *J Manipulative Physiol Ther* 1998;21:1–7.
29. Joyce CR, Zutshi DW, Hrubes V, et al. Comparison of fixed interval and visual analogue scales for rating chronic pain. *Eur J Clin Pharmacol* 1975;8:415–20.
30. Breivik EK, Björnsson GA, Skovlund E. A comparison of pain rating scales by sampling from clinical trial data. *J Clin Pain* 2000;16:22–8.
31. van Laerhoven H, van der Zaag-Loonen HJ, Derkx BH. A comparison of Likert scale and visual analogue scales as response options in children's questionnaires. *Acta Paediatr* 2004;93:830–5.
32. McQuay H. Consensus on outcome measures for chronic pain trials. *Pain* 2005;113:1–2.
33. Price DD, McGrath PA, Rafii A, et al. The validation of visual analogue scales as ratio scale measures for chronic and experimental pain. *Pain* 1983;17:45–56.
34. Price DD, Bush FM, Long S, et al. A comparison of pain measurement characteristics of mechanical visual analogue and simple numerical rating scales. *Pain* 1994;56:217–26.
35. Maxwell C. Sensitivity and accuracy of the visual analogue scale: a psychophysical classroom experiment. *Br J Clin Pharmacol* 1978;6:15–24.
36. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960;20:37–46.
37. Downie WW, Leatham PA, Rhind VM, et al. Studies with pain rating scales. *Ann Rheum Dis* 1978;37:378–81.
38. Collins SL, Moore RA, McQuay HJ. The visual analogue pain intensity scale: what is moderate pain in millimeters? *Pain* 1997;72:95–7.