OsteoArthritis and Cartilage (2004) **12,** 225–231

doi:10.1016/j.joca.2003.11.006

Osteoarthritis and Cartilage

International Cartilage Repair Society

OARSI OSTEOARTHRITIS RESEARCH SOCIETY INTERNATIONAL

Relative efficiency and validity properties of a visual analogue vs a categorical scaled version of the Western Ontario and McMaster Universities Osteoarthritis (WOMAC) Index: Spanish versions

I. Villanueva MD†*, M. del Mar Guzman MD†, F. Javier Toyos MD‡, R. Ariza-Ariza MD‡ and F. Navarro MD‡

† *University of Arizona Arthritis Center, Tucson, AZ, USA*

‡ *Hospital Universitario Virgen Macarena, Sevilla, Spain*

Summary

Objective: To compare the performance of visual analogue (VAS) vs categorical (CT) scaled versions of the three subscales (pain, stiffness and difficulty) of the WOMAC Index in patients with knee osteoarthritis.

Material and methods: Patients with knee OA (at least grade II of Kellgren & Lawrence classification) were treated for a 6-weeks period with an NSAID. The following measures were applied at baseline and after treatment: VAS and CT WOMAC scales, Lequesne Index, pain on VAS, and global assessment by patient and observer. Construct Validity was determined by correlation of each of the subscale scores with the other outcomes both at baseline and after treatment (Pearson's test); responsiveness comparing baseline vs final status by Wilcoxon's test; reliability by analysis of the internal consistency using Cronbach's alpha at baseline and after the treatment period; and test–retest reliability by Kendall's Tau-c statistics. Finally, correlation and analysis of the relative efficiency (RE) between the subscales of both formats were tested.

Results: Seventy-three (8 men and 65 women) patients with a mean disease duration from first diagnosis of 69 months (3 to 254) completed the study. The following were the observed values for the instruments' psychometric properties: (1) construct validity: correlations ranged from 0.30 to 0.84 for VAS and 0.27 to 0.77 for CT subscales; (2) responsiveness: achieved *P* values for the pain, stiffness and difficulty scales were $P<0.0001$, $P=0.002$ and $P<0.0001$ in VAS and $P=0.003$, $P<0.0001$ and $P=0.001$ in CT format respectively; (3) internal consistency: the obtained Cronbach's alpha coefficients ranged from 0.71 to 0.97 for the VAS and 0.64 to 0.95 for the CT subscales; (4) test–retest reliability: correlation coefficients ranged from 0.36 to 0.76 for VAS and 0.34 to 0.52 for CT subscales; and (5) the relative efficiency of the subscales in VAS vs CT format were 2.20, 0.91 and 1.91 for pain, stiffness and difficulty respectively. Significant correlations between subscales in both formats ranging from 0.72 to 0.86 were observed.

Conclusion: We have shown that both the VAS version of WOMAC have adequate evaluative and discriminative properties. We found the pain and physical function scales in VAS format and the stiffness scale in CT format to have a slightly better performance in this sample.

Key words: WOMAC Index, Validation, Knee osteoarthritis, Outcomes, Clinical trials.

Introduction

The Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) is a disease and joint specific instrument, developed for the evaluation of knee and/or hip osteoarthritis (OA). Since its initial validation it has been widely used and has become the measure of choice for the assessment of OA patients both in clinical trials and observational studies. The questionnaire includes three subscales that target three of the most relevant outcomes in OA[1]: specifically, the dimension of pain (5 items), stiffness (2 items) and physical function (17 items).

The use of standard and validated instruments that have been subject of a sound process of transcultural adaptation and revalidation is critical when the researchers' goal is to perform international – even inter-regional – comparative analyses of disease outcomes, to carry out analytic systematic reviews, or to conduct multinational clinical trials[2–4]. Moreover, when performing outcomes evaluation, especially in the setting of clinical trials, the use of instruments with the highest established levels of reliability and discriminative power becomes crucial. Otherwise the statistical estimates that are obtained constitute highly attenuated versions of the true population parameters, which could have a great impact on the required sample size and, hence, on the budget required for the study. It is suggested that, when giving their answers to the formulated questions, some individuals find it difficult to place their responses on visual analogue scales (VAS) and prefer to do so on categorical (CT) scaled questionnaires[5,6]. This could further increase measurement error and have an additional impact on the reliability of that measurement. The WOMAC Index is available in both VAS and CT formats, but, to our best knowledge, only one comparative evaluation of both

*Address reprint requests to: Dr Isidro Villanueva-Torrecillas, Assistant Professor, Arizona Arthritis Center, University Medical Center, Room 7341A, 1501 N Campbell Ave, Tucson, AZ 85724-5093, USA. Tel.: +1-520-626-6046; +1-520-626-5018; E-mail: isidro@u.arizona.edu

Received 7 April 2003; revision accepted 2 November 2003.

formats of the questionnaire has been published, which was performed in the context of the original validation of the instrument[7–9]. More recently, the development and validation of a computerized version of the Index has been reported[10]. We have reported elsewhere the transcultural adaptation and re validation of Spanish visual analogue and categorical scaled versions of the WOMAC Index[11]. The objective of the present study was to comparatively evaluate the performance of visual analogue vs categorical scaled versions of the three WOMAC Index subscales (pain, stiffness and difficulty) in a sample of patients with knee osteoarthritis that participated in a NSAID clinical trial in Spain.

Subjects and methods

Patients with symptomatic OA of the knee, as defined by the American College of Rheumatology (ACR) criteria[12], were screened and recruited to participate in a multi-center, double-blind, randomized, clinical trial for the comparison of two formulations of the same nonsteroidal anti-inflammatory drug (NSAIDs). Consecutive patients attending the participating rheumatology clinics from November 1997 to December 1998 were screened to participate in the study. Patients were recruited after obtaining informed consent and verifying the inclusion and exclusion criteria (as defined per clinical trial protocol). Approval from the corresponding institution's review board was obtained. Data of subjects recruited in two of the centers was used for the present analysis, which was conducted at the Department of Rheumatology, Virgen Macarena University Hospital, Seville, Spain.

Inclusion criteria were: age of 35–75 years; a disease duration (from diagnosis of knee OA) of at least one year; knee pain on most days in the last month, with a score of at least 40 mm on a 0 to 100 mm VAS; persistent knee pain after a 2-week NSAID wash-out period; a score of 4 to 14 (both inclusive) on the Lequesne Index (see description below); and definite radiographic evidence of primary knee OA, defined as at least grade 2 (Kellgren & Lawrence classification[13]) on an X-ray performed in the previous 6 months. Potential participants were excluded if they had: inflammatory arthritis (e.g., rheumatoid arthritis, psoriatic arthritis) any comorbidity that precluded safe use of NSAIDs; and, any condition that interfered with outcomes evaluation.

Assessments were performed at the outpatients' clinics in each center. Data for this study corresponded to the baseline and post-treatment (6 weeks) evaluations. Data was collected in an interview format using the below described self-reported measures and instruments.

MEASUREMENTS

The *Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)*[7–9]. This is a validated, 24-item, disease and joint-specific measure that evaluates knee pain, stiffness and physical function. The instrument has been designed as a self-report measure. The time frame for evaluation selected for this study was the previous 72 h. As said, the instrument contains three subscales:

1 The physical function (difficulty) subscale comprises 17 items. Participants are asked to rate the intensity of their difficulty to perform a wide range of activities, such as ascending or descending stairs, walking, bending or performing daily activities, on a 0 to 100 horizontal scale anchored with the Spanish words corresponding to 'No difficulty' and 'Very intense difficulty' on the left and right extremes respectively. Bellamy *et al.* showed the adequate validity (correlation coefficients ranging from 0.36 to 0.59) excellent internal consistency (Cronbach's alpha 0.91) and test–retest reliability (Kendall Tau-c coefficient=0.72) of this scale[9].
2 The pain subscale comprises five items. Patients were asked about the intensity of pain in the reference knee. Activity pain, night pain and pain at rest were assessed with the five items that constitute this subscale. A 0 to 100 VAS format was used, as described for the physical function subscale, for each of the items; in this case anchored with the Spanish words corresponding to 'No pain' and 'Very intense pain'. This scale has also shown adequate validity (correlation coefficients ranging from 0.40 to 0.62), high internal consistency (Cronbach's alpha 0.81) and test–retest reliability (Kendall Tau-c coefficient=0.68)[9].
3 The stiffness subscale of the WOMAC Index comprises two items. Participants were asked about the intensity of the stiffness in two different situations: during the morning, right after waking up, and during the day after some period of rest or inactivity. The 0 to 100 VAS was anchored with the Spanish words corresponding to 'No stiffness' and 'Very intense stiffness'. This scale has shown adequate validity (correlation coefficients ranging from 0.32 to 0.46) excellent internal consistency (Cronbach's alpha 0.90) and test–retest reliability (Kendall Tau-c coefficient=0.61)[9].

The WOMAC Index was also applied in five-point categorical-answer format to meet the objective of the comparative performance of both formats. The adjectives located at the extreme of the categorical scales were the same that were used to anchor the above-described VA scales, plus the adjectives 'mild', 'moderate', and 'intense' were added to label the five categories.

Lequesne algo-functional Index[14,15]

This index has been used extensively in European OA studies since 1980. The index includes three sections with a total of ten questions and takes a few minutes to complete. The first section asks about pain or discomfort 'At night', 'After getting up in the morning', 'When standing', 'When walking' and 'When rising from sitting' (knee index) and pain when 'Sitting two hours' (hip index). The items are graded with dichotomous or categorical answers. The second section asks about the maximum walking distance. If patients use one or two walking aids (canes, crutches. . .) the score is raised by one or two points, respectively. The third section addresses physical function disability with five categories, graded from 0= 'without difficulty' to 2='unable to do'. The Lequesne OA-index is scored as the sum of all questions. The score range of each section is from 0 to 8 resulting in a total score ranging from 0 to 24. The Lequesne OA-index waseen developed using an interview format.

Global knee pain on a 0 to 100 VAS, answering to the question: How much pain have you had in your knee during the past week? Participants were instructed to chose a reference knee. The scale was anchored with the words 'No pain' and 'Most severe pain possible' at the left and right extreme, respectively. This type of scale has shown

high levels of reliability (coefficients ~0.90)[16] and validity (correlation coefficients with other scales measuring pain ranging from 0.71 to 0.91)[17].

Patient and physician's global assessment of disease activity

Participants were asked to rate the activity of the arthritis on the reference knee in the previous week on a 0 to 100 VAS anchored with 'No activity' and 'Very severe activity'. Physicians were required to complete this assessment after physical examination and blinded from the results of patient's evaluations.

Patient's perceived level of improvement

At the 6-week evaluation, patients were asked to report their perceived level of improvement on a seven-point Likert-type scale that included the following adjectives: much worse, worse, somewhat worse, the same, somewhat better, better, much better. This scale was used to identify patients that remained clinically stable after the six weeks of treatment.

CROSS-CULTURAL ADAPTATION

The two versions of the index (VAS and CT formats) were adapted following standard guidelines for cross-cultural adaptation of health-related quality of life measures[3]. The results of this adaptation has been reported elsewhere[11]. In brief, three bilingual persons (two Spanish investigators and a native English speaking teacher) performed a translation of the original instrument. The three versions were reviewed and, due to the similarity of all of them, it was decided to obtain only one back translation to be performed by the teacher, who was unaware of the intent of the original instrument. Then, a committee composed of three investigators (rheumatologists) and a research nurse reviewed the source and the final versions. The task of the committee was to identify potential discrepancies and problems with specific items and solve them to ensure that the final version maintained content validity and conceptual equivalency. Few items were modified during this process. In general, it was perceived that, due to the characteristics of the original instrument, the process of adaptation was straightforward; source and final versions were considered to be equivalent. Only one item corresponding to the difficulty subscale was problematic: 'Bending to floor'. The discrepancy was resolved after consulting with a native English speaking person. Finally, the final instrument was pre-tested on a sample of 10 OA patients. Subjects were asked if they found any difficulty understanding the instructions and the items included in the questionnaire. None of them expressed any difficulty understanding what they were asked in any of the items. The only source of difficulty that we found was related with the format of the answer itself. Some subjects showed initial problems in understanding the process of answering on a VAS, but that difficulty was overcome after a short training with example questions. The interviewer exemplified marking on the scale the potential responses in the hypothetical questions/scenarios were the patient would have 'no pain at all', 'average/moderate pain' and 'pain very intense'. For example, the interviewer showed to the patients where they would put their marks on the VAS if they did not have any pain (just on the left anchor in this case).

ANALYSIS

Following standard guidelines for the evaluation of measurement properties of quality of life instruments[18], we comparatively tested the validity, reliability and responsiveness of both Spanish versions (VA and CT) of the WOMAC Index. The feasibility of each instrument was also evaluated as a measure of their potential applicability in the Spanish population. This was carried out by assessing the number of subjects that completed the questionnaire; their level of comprehension of the instruments' content on a five-point categorical scale; and the required time for completion of the instrument.

Validity

We examined construct validity (convergent and divergent – or discriminant) by correlating the scores of the index subscales with the other measures applied in the study (see measurements). A particular subscale is expected to converge with the scores of those instruments targeting the same construct, and to deviate from the scores given by instruments or scales assessing a different one. To quantify these relationships, Pearson's correlation coefficients were obtained.

Reliability

It was examined in two different ways. First, we assessed the internal consistency, which gives a measure of the stability across the items that constitute a particular scale and with the total score of the scale. Internal consistency was tested using Cronbach's alpha coefficients, which were obtained at baseline and post-treatment. Second, we examined the test–retest reliability, which examines the stability of the measurements across evaluations with the same scale performed at different time points. To test this property, Intraclass Correlation Coefficients were obtained for those patients with unchanged clinical status after treatment, as perceived by the patient.

Responsiveness

We tested the sensitivity to change of both versions of the index by assessing the ability of their corresponding subscales to detect within-subject changes. This was achieved by comparing the scores at baseline and 6-weeks (post-treatment) using a Wilcoxon's Rank test. The responsiveness of both versions of the instrument was determined by computing their relative efficiency, This estimtes the extent to which one scale is more or less efficient at detecting change over time relative to another scale. Relative efficiency and the significance of change scores are typically computed from paired-samples *t*-tests as pairwise squared *t*-values (t^2 scale I/t^2 scale 2).[19] However, results generated by this method confound 'responsiveness' with the effects of nonnormality such that scales with more normally distributed outcomes are favored. Therefore, we examined the significance of chance scores using Wilcoxon's matched-pairs signed-ranks test and computed relative efficiency as pairwise squared *z*-values z^2 scale I/z^2 scale 2) generated by this test[20]. The z-value for the WOMAC scales in VAS format were chosen as the denominators so that values for the WOMAC scales in CT format estimate their relative efficiency as a percentage of the VAS scales.

Table I
Characteristics of participants at baseline

	n=73
Age (mean-range)	58.71 (35–73)
Sex (male/female)	8/65
Mean osteoarthritis duration (months from diagnosis)	77.62 (3–264)
Educational level (*n*, %)	
Some primary school	62(84.9%)
Complete primary school	7(9.6%)
High school	3(4.1%)
College	1(1.4%)
VAS – WOMAC subscales*	
Pain – VA (0 to 50)	26.821(7.65)
Stiffness – VA (0 to 20)	11.74(4.02)
Physical function – VA. (0 to 170)	86.87(28.03)
CT – WOMAC subscales*	
Pain – CT (0 to 20)	10.38(3.51)
Stiffness – CT (0 to 8)	4.93(1.69)
Physical function – CT (0 to 68)	34.16(11.51)
Lequesne Index scores*	
Pain (0 to 8)	5.01(0.89)
Physical function (distance+activities) (0 to 16)	5.47(1.65)
Total score (0 to 24)	10.47(2.06)
Pain on VAS (0 to 100)	66.68(20.81)
Global evaluation by the patient on VAS (0 to 100)*	68.01(15.39)
Global evaluation by the observer on VAS (0 to 100)*	58.67(14.18)
Radiological classification, *n* (%) (Kellgren & Lawrence)	
Grade II	14(19.40)
Grade III	45(61.20)
Grade IV	14(19.40)

*Mean (S.D.)

Results

Data from 73 (eight men and 65 women) subjects with a confirmed diagnosis of knee OA, and meeting the trial's inclusion and exclusion criteria, were used for this analysis. The average disease duration from OA diagnosis was 74 months (range 3 to 264) and the mean patients' age was 57.3 years (range 35 to 73). Patients' characteristics at baseline are shown in Table I.

FEASIBILITY

The average time for completion of the three subscales of the WOMAC Index was 7.24 and 7.48 minutes for the VAS and CT formats respectively. All participants completed the instruments both at baseline and 6-weeks evaluations. To examine participants' level of comprehension of the instruments' content, a proxy question was asked; did you have any difficulty understanding the questionnaire items? (To be answered on a five-point CT scale). Sixty-one participants affirmed they had 'no difficulty', ten participants found 'some difficulty' and only two respondents seemed to have 'moderate difficulty' in understanding and responding to the items.

DISCRIMINATIVE PROPERTIES

Validity

Construct validity. As hypothesized, several positive and highly significant correlation coefficients (Pearson's test) between the subscales corresponding to both the VAS and CT versions of the WOMAC Index and Lequesne subscales, global pain on 0 to 100 VAS, and Global assessments by the patient and the physician, were obtained. Correlation coefficients ranged from 0.30 to 0.84 for the VAS subscales and 0.27 to 0.77 for the CT versions. All subscales showed a good level of convergence with other measures testing the same dimension or construct, as demonstrated by the observed highly significant correlations. Table II shows the coefficients for the VAS and CT versions, respectively. Both versions also showed a sufficient level of discriminant validity (divergent construct validity), as demonstrated by the fact that the correlation coefficients were higher with measures testing the same dimension than with measures testing a different construct (Table II). Similar coefficients were obtained for both the VAS and CT versions, indicating very similar levels of construct validity.

Finally, high correlation coefficients were obtained between VAS and CT pain, stiffness and physical function subscales both at baseline (Table III: the nine values in the left upper section) and final evaluation (Table III: the nine values in the right lower section). In general, the coefficients were higher between subscales measuring the same construct and lower between those targeting a different dimension. The coefficients were in general higher for the final evaluation, which probably indicates a decrease in measurement error as a result of a learning effect from the baseline assessment. These results further support the convergence and discriminative ability of the subscales in both formats, and indicate a high level of agreement between both Spanish versions of the WOMAC Index.

Criterion validity. We also tested the discriminative properties of the two versions of the WOMAC Index by contrasting their respective total score with the dichotomized patient's global evaluation, which was measured on a 0 to 100 VAS; the dichotomization was based on the observed median score, which split the sample in two groups: subjects with worse vs better perceived disease status. To examine the ability of the instruments to discriminate subjects in each group, ROC curves plotting the sensitivity vs 1-specificity were obtained. The observed areas under the curve were 0.75 (95% CI, 0.64–0.86) and 0.68 (95% CI, 0.55–0.80) for the VAS and CT versions, respectively, indicating only slightly better discriminative ability for the VAS format (Fig. 1).

Reliability

Internal consistency. The Cronbach's alpha coefficients obtained both at baseline and final evaluations for the scales in VAS and CT format are shown in Table IV. In general the coefficients were slightly higher for the VAS subscales and better in the second evaluation, which again probably indicates a decrease in measurement error as a consequence of some learning after the first evaluation. The obtained coefficients for the post-treatment evaluation

Table II
Correlation coefficients for the WOMAC–VAS/CT scales and other measures (Pearson's Test)

WOMAC scales	Lequesne pain	Lequesne function	Lequesne stiffness	Global pain (VAS)	PGA (VAS)	OGA (VAS)
Pain – VAS	0.38*	0.53**	0.38*	0.63**	0.42**	0.62**
Stiffness – VAS	0.08	0.38*	0.70**	0.43**	0.37*	0.56**
Physical function – VAS	0.33†	0.72**	0.61**	0.68**	0.43**	0.78**
Pain – CT	0.31†	0.52**	0.57**	0.53**	0.33‡	0.64**
Stiffness – CT	0.17	0.46**	0.71**	0.49**	0.31‡	0.57**
Physical function – CT	0.24†	0.66**	0.58**	0.62**	0.37*	0.70**

* P=0.001.
** P<0.001.
† P<0.05.
‡ P<0.05.
VAS=visual analogue scale; CT=categorical; PGA=patient global assessment; OGA=observer global assessment.

Table III
Correlation coefficients for the WOMAC–VAS and CT scales at both baseline and final evaluations (Pearson's Test)

	Pain VAS (b)*	Stiffness VAS (b) *	Physical function VAS (b) *	Pain VAS (f)†	Stiffness VAS (f) †	Physical function VAS (f) †
Pain – CT (b) *	0.72	0.39	0.83			
Stiffness – CT (b) *	0.70	0.75	0.82			
Physical – CT (b) *	0.68	0.55	0.86			
Pain – CT (f) †				0.57	0.70	0.68
Stiffness – CT (f)				0.74	0.81	0.75
Physical – CT (f) †				0.71	0.75	0.86

*(b)=Baseline evaluation.
†(f)=Evaluation after 6 weeks of treatment. VAS=visual analogue scale; CT=categorical.

were excellent for both versions of the Index with somewhat smaller values for the stiffness subscale, which was expected for a scale with only two items.

Test–retest reliability. It was evaluated only in the 17 patients that reported remaining clinically stable during the study period (answered they were 'the same' on the cited seven-point Likert-type scale). The obtained Intraclass Correlation Coefficients between the baseline and post-treatment evaluations for both the scales in VAS and CT format are shown in Table IV. Despite the strategy used to determine test–retest reliability and the time difference between both evaluations, the obtained coefficients were more than acceptable. Again, the coefficients were smaller for the stiffness subscale in both the VAS and CT versions, probably due to the fact that they are composed only by two items. Test–retest reliability was also tested in those same patients by comparison of the baseline and post-treatment evaluations using a Wilcoxon's Rank test. As desired, the comparisons for each subscale yielded a non-significant result (P values ranging from 0.12 to 0.88 for the VAS version, and 0.12 to 0.85 for the CT version).

Evaluative properties

Responsiveness. The ability to detect change of the scales was evaluated by means of the comparison of the baseline and post-treatment evaluations, using Wilcoxon's Rank tests. The results of these comparisons are shown in Table V. A highly statistically significant improvement was observed for all WOMAC subscales both in VAS and CT format after the treatment period. Based on the obtained P values, the pain and physical function subscales of the WOMAC Index in CT format were slightly less sensitive to change than the VAS version, but the stiffness subscale was slightly more sensitive in CT format. The scales of Global Knee pain, Global assessments by the patient and the physician, Lequesne Pain and Lequesne Physical Function also improved significantly (data not shown).

Relative efficiency statistics showed that the pain and physical function scales in CT format were estimated to be 48% and 52% as efficient at detecting change as the same scales in VAS format in this sample (relative efficiency coefficients: 0.48 and 0.52, respectively). However, the stiffness scale in CT format proved to be 21% more efficient than the scale in VAS format at detecting change (relative efficiency coefficients: 1.21).

Discusion

The results of this study suggest that both the VAS and CT Spanish versions of the WOMAC have adequate evaluative and discriminative properties. We have shown both versions of the Index to be feasible, to have excellent levels of reliability and validity, and to be highly responsive to within-subject changes over time. These results were very similar to those reported in the original validations of the English versions of the WOMAC Index[7,8]. Similar to Bellamy *et al.*[7,8], the evaluations were performed in the context of a clinical trial. Although the evaluation context could differ in other types of studies, there is no reason to believe that the instruments would show worse performance in the setting of observational studies or even in the routine clinical practice[9]. The observed measurement properties clearly support their use in the evaluation of pain, stiffness, and physical function levels in patients with knee OA, which were to be recruited from the general Spanish population.

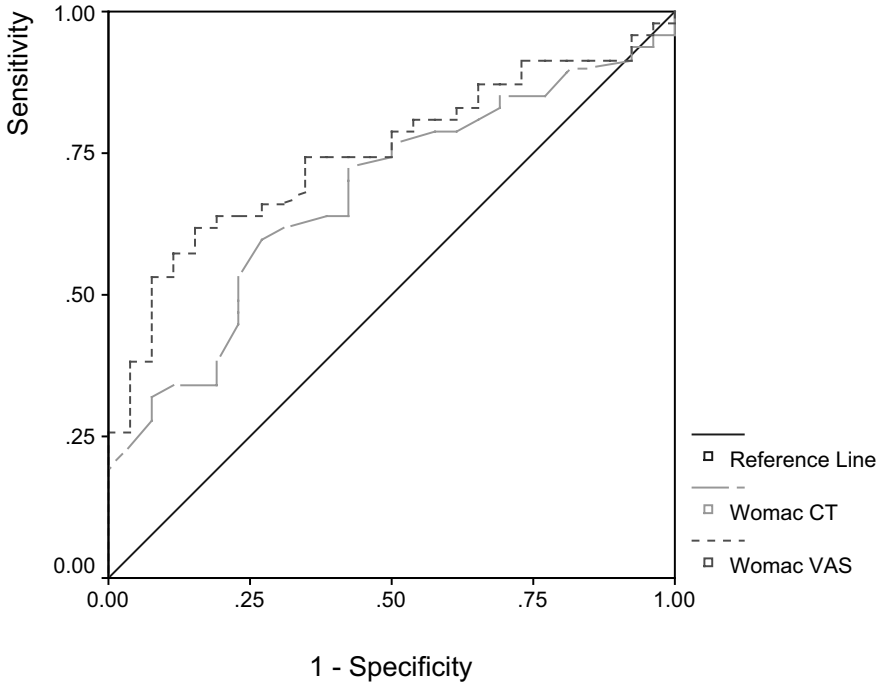

Fig. 1. ROC curves for criterion validity test.

Although both versions of the questionnaire showed good performance, similar measurement properties and a fairly good level of agreement – as reported by the researchers that developed the Index[9], we found the VAS version of the WOMAC Index to have slightly higher levels of convergent and discriminant validity, internal consistency, test–retest reliability and responsiveness. However, the stiffness subscale in CT format proved to be more reliable, as evidenced by its slightly higher levels of both internal consistency and test–retest reliability, than the corresponding VAS version. The relative efficiency was, in general, higher for the scales in VAS format as compared to those in CT format.

Table IV
Reliability coefficients: Internal consistency (Cronbach's AA) and Test-retest (Intraclass correlation coefficient)

WOMAC scales	Internal consistency		Test–retest
	Baseline	Final	
Pain – VAS	0.71	0.91	0.71
Stiffness – VAS	0.74	0.81	0.50
Physical function – VAS	0.94	0.97	0.83
Pain – CT	0.77	0.79	0.55
Stiffness – CT	0.64	0.76	0.62
Physical function – CT	0.93	0.95	0.77

Table V
*Responsiveness of the WOMAC subscales (Wilcoxon's Rank Test)**

	Baseline	Final	*P*
Pain – VAS	26.82 (7.65)	20.82 (10.82)	<0.0001
Stiffness – VAS	11.74 (4.02)	8.63 (5.17)	0.002
Difficulty – VAS	97.38 (28.02)	78.25 (34.01)	<0.0001
Pain – CT	10.38 (3.51)	8.60 (3.48)	0.003
Stiffness – CT	4.93 (1.69)	3.81 (1.92)	<0.0001
Difficulty – CT	38.9 (11.50)	33.05 (12.57)	0.001

*Results are expressed in mean (S.D.).
VAS=visual analogue scale; CT=categorical.

These results are consistent with what was reported in the original validation study, where the VAS version of the Index was found to be more sensitive to changes[7]. Those authors suggested in their report that the selection of one or the other version of the Index could depend on researchers' preferences. From our results we would recommend the use of the VAS version, but we recognize that the CT version of the Index is a good alternative in the case that either the researcher or the patient should prefer this format. We think that the CT version would benefit from changing the number of possible response categories from 5 to 7; this would probably enhance its measurement properties, particularly its sensitivity to change, but it would require a new validation study to confirm this possibility. We recognize that some of the observed differences between VAS and CT versions could reside in the intrinsic characteristics of the computation of scores and the resulting frequency distributions for the obtained values. To explore this potential source of error we performed the analyses with both parametric and non-parametric tests, which essentially gave the same results.

The three subscales of the Index showed similar measurement properties – both in VAS and CT format but the stiffness subscale seemed to have worse performance. This could reside in the fact that this particular scale is the shorter (i.e., only composed of two items). However we found the stiffness subscale to be more reliable in CT format, which could be due, at least in part, to the recognized difficulty that patients have in expressing their level of perceived stiffness[10]. In this case, patients probably found it easier to score their stiffness on a scale with concrete descriptions of the intensity level of stiffness.

Some investigators have reported that patients find it difficult to respond to VAS scaled questionnaires, especially illiterate patients[5,6]. We have found that not to be the case in our sample, which was composed of subjects with low educational level. Our results suggest that subjects exposed to a single previous application of the instrument increase their confidence with the measurement in VAS format, as shown by the higher internal consistency levels observed in the final evaluation. The evaluations were in an interview format, which could have some influence on the ability of subjects to respond on VAS. The developers of the WOMAC Index found the same in their research[10].

In conclusion, we have shown that both the VAS and CT Spanish versions of the WOMAC have adequate evaluative and discriminative properties. We found the pain and physical function scales in VAS format and the stiffness scale in CT format to have slightly better performance in this sample, which could have some impact on the required sample size to detect meaningful differences.

References

1. Bellamy N. Osteoarthritis Clinical Trials: Candidate variables and clinimetric properties. J Rheumatol 1997;24:768–78.
2. Bellamy N, Buchanan WW. Outcome measurement in osteoarthritis clinical trials: The case for standardisation. Clin Rheumatol 1984;3:293–303.
3. Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health related quality of life measures: literature review and proposed guidelines. J Clin Epidemiol 1993;12:1417–32.
4. Anderson GF. Analyzing health outcome through international comparison. Med Care 1994;32:526–34.
5. Huskisson EC. Measurement of pain. Lancet 1974; 2:1127–31.
6. Bosi Ferraz M, Quaresma MR, Aquino LRL, Atra E, Tugwell P, Goldsmith CH. Reliability of pain scales in the assessment of literate and illiterate patients with rheumatoid arthritis. J Rheumatol 1990;17:1022–4.
7. Bellamy N, Buchanan WW, Goldsmith CH, *et al.* Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol 1988;15:1833–40.
8. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt L. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes following total hip or knee arthroplasty in osteoarthritis. J Orthop Rheumatol 1988; 1:95–108.
9. Bellamy N. WOMAC Osteoarthritis index – a user's guide. London: London Health Sciences Center, 1998.
10. Bellamy N, Campbell J, Stevens J, Stewart C, Mahmood Z. Validation study of a computerized version of the Western Ontario and McMaster Universities VA 3.0 Osteoarthritis Index. J Rheumatol 1997; 24:2413–5.
11. Villanueva I, Toyos FJ, Ariza R, Guzman MM, Navarro F. Validation of a spanish version of the WOMAC (Western Ontario and McMaster Universities osteoarthritis Index). Arthritis Rheum 1997;40:S118.
12. Altman R, Asch E, Bloch D, *et al.* Development criteria for the classification and reporting of osteoarthritis. Classification of Osteoarthritis of the knee. Diagnostic and Therapeutic Criteria Committee of the American Rheumatism Association. Arthritis Rheum 1986;29:1039–49.
13. Kellgren JH, Lawrence JS. Radiological assessment of osteoarthrosis. Ann Rheum Dis 1957;16:494–501.
14. Lequesne MG, Mery C, Samson M, Gerard P. Indexes of severity for osteoarthritis of the hip and knee. Validation value in comparison with other assessment tests. Scand J Rheumatol 1987;65:85–9.
15. Lequesne MG. The algofunctional indices for hip and knee osteoarthritis. J Rheumatol 1997;24:779–81.
16. Dubuisson D, Melzack R. Classification of clinical pain descriptions by multiple group discriminant analysis. Exp Neurol 1976;51:480–7.
17. Downie WW, Leatham PA, Rhind VM, *et al.* Studies with pain rating scales. Ann Rheum Dis 1978; 37:378–81.
18. Guyat GH, Feeny DH, Patrick DL. Meausring health-related quality of life: basic science review. Annals Int Med 1993;70:225–30.
19. Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures: statistics and strategies for evaluation. Control Clin Trials 1991;12:142S–158S.
20. Hobart JC, Riazi A, Lamping DL, Fitzpatrick R, Thompson AJ. Meausring the impact of MS on walking ability: the 12-Item MS Walking Scale (MSWS-12). Neurology 2003;60:31–6.