

Clinimetric properties of WOMAC index in Greek knee osteoarthritis patients: comparisons with both self-reported and physical performance measures

George Papathanasiou · Sophia Stasi · Leon Oikonomou ·
Ioanna Roussou · Effie Papageorgiou · Efstathios Chronopoulos ·
Nectarios Korres · Nicholas Bellamy

Received: 26 November 2013 / Accepted: 12 May 2014 / Published online: 29 May 2014
© Springer-Verlag Berlin Heidelberg 2014

Abstract This observational study aimed to examine the clinimetric properties of the Greek for Greece translation of the Western Ontario and McMaster Osteoarthritis Index (WOMAC®). One hundred and twenty-three patients with knee osteoarthritis (mean age 69.5 years) participated in the study. An extensive reliability study was carried out to assess WOMAC's internal consistency and repeatability (8-day interval). In addition, we examined the construct (convergent, nomological and known-groups) and criterion-related (concurrent and predictive) validity of the index against both self-report [SF-36 and combined visual analog/faces pain scale-revised (VAS/FPS-R)] and physical performance measures [timed up and go test (TUG)]. The internal consistency of the WOMAC subscales ranged from high (0.804) to excellent (0.956). Intra-class correlation coefficients for test–retest reliability were excellent, ranging from 0.91 to 0.95. Partial correlation analysis, adjusted for age and use of an assistive device, showed that WOMAC scores were significantly associated with

all validation criteria, presenting fair to strong (−0.33 to −0.86) correlation coefficients. WOMAC-function was strongly associated with SF36-function (−0.86) and TUG (0.71), WOMAC-pain to VAS/FPS-R (0.71) and SF36-pain (−0.67). Of all WOMAC outcomes, stiffness subscale had the lowest, though still significant, correlations with all validation criteria. Multiple linear regression analyses indicated that WOMAC-function was a significant factor for TUG, WOMAC-pain for VAS/FPS-R and both for SF36-function and SF36-pain. The WOMAC LK3.1 Greek for Greece Index is a reliable and valid assessment tool for the evaluation of individuals with knee osteoarthritis, showing excellent reliability and significant validity properties.

Keywords Knee osteoarthritis · SF-36 · TUG test · VAS scale · WOMAC® index

Introduction

Osteoarthritis (OA) is the most common type of degenerative joint disease and is a leading cause of chronic disability [1]. The disease most commonly affects the middle-aged

Electronic supplementary material The online version of this article (doi:10.1007/s00296-014-3043-x) contains supplementary material, which is available to authorized users.

G. Papathanasiou (✉) · S. Stasi · I. Roussou
Physiotherapy Department, Technological Educational Institute
of Athens, 24 Mitrodorou St., 10441 Athens, Greece
e-mail: papathanasiou.g@gmail.com

L. Oikonomou
Resident of A' Orthopaedic Department, General Hospital
of Athens "G. Gennimatas", 154 Mesogion Ave., 11527 Athens,
Greece

E. Papageorgiou
General Department of Mathematics, Technological Educational
Institute of Athens, Ag. Spyridonos St., 12210 Athens, Greece

E. Chronopoulos
2nd University Orthopaedic Department, Konstantopoulou
Hospital, 3-5 Agias Olgas St., 14388 N. Ionia, Athens, Greece

N. Korres
1st Orthopaedic Department, General Hospital of Attica "KAT",
2 Nikis str., 14561 Kifissia, Athens, Greece

N. Bellamy
School of Medicine, Royal Brisbane and Women's Hospital, The
University of Queensland, Edith Cavell Building (Room 127),
Herston, QLD 4029, Australia

and elderly, although younger people may also be affected as a result of injury or fatigue. Knee and hip osteoarthritis are leading causes of lower extremity pain and disability in the general population [2], with knee OA being the fourth most important cause of disability in women and the eighth in men.

The growing demand for evidence-based practice requires the utilization of reliable and validated instruments as evaluation tools, so as to provide clinicians with objective and quantified health status data. In large-scale studies, specially designed self-reported questionnaires and rating scales are frequently used for the functional assessment of OA patients; such instruments have the advantages of being noninvasive, quick to complete, and easy to administer. One such disease-specific evaluation tool is the Western Ontario and McMaster Index (WOMAC[®]), developed in the early 1980s for measuring the level of pain, joint stiffness, and functional ability in patients with hip or knee OA [3, 4]. Multinational studies have shown that the WOMAC index has strong clinimetric properties, being the most widely used measure for assessing self-reported pain, stiffness, and function in patients with hip or knee OA.

In recent years, the importance of functional enhancement and pain management in the treatment of OA has acquired greater prominence. For this reason, a thorough assessment of a patients' pain and functional level is an essential prerequisite for the choice of the most effective treatment plan and contributes to the decision-making process for surgical or non-surgical intervention. Self-report instruments, rather than physical performance tests, represent the preferred method of assessing the health concepts of pain and physical function in OA patients [5, 6]. In addition, self-report instruments have been favored over physical performance measures as external criteria in most WOMAC studies that examined the validity of physical function and pain domains in different population groups [7–9]. However, it has been shown that these evaluation tools offer complementary information [9]; thus, both are needed to perceive the multidimensional impact of pain [6] and capture the construct of physical function in its entirety [8].

The purpose of the present study was to examine the clinimetric properties of the WOMAC index in Greek individuals with knee OA, given that the prevalence of knee OA in Greece is many times greater than that of OA in other major joints [10]. Examination of the reliability and validity properties of the cross-culturally adapted Greek version of the WOMAC index would allow its broader clinical use in hip or knee OA patients and may add to the overall clinimetric properties of the instrument. In addition, we sought to extend the study of the validity properties of the WOMAC index by testing it against both self-report and physical performance measures. A broader

awareness of these findings in the Greek setting would facilitate objective comparisons between studies of different national origin and could contribute to the validity of future meta-analyses.

Methods

Study population

One hundred and fifty community-dwelling men and women, aged 65 years and over, were randomly selected from five municipal "Open Care Centers for the Elderly" and invited to participate in this observational cross-sectional study. The main inclusion criterion was the existence of knee OA, according to the American Rheumatism Association's (ARA) functional class I, II, or III criteria [11]. In order to be eligible, patients had to report pain on motion, have a clinical and radiographic diagnosis of knee OA, and have been using anti-inflammatory medication and/or receiving physical therapy for at least the previous 6 months [12]. Participants were excluded if they had undergone any kind of surgical intervention to the affected knee; had medical conditions such as rheumatoid arthritis, psoriatic arthritis, systemic lupus erythematosus, lower limb muscle weakness due to a central or peripheral neurological etiology, unstable angina, or uncontrolled hypertension or hypotension; or were taking medication that adversely affected their postural or dynamic balance [13]. Written informed consent was obtained from all participants. The study protocol was approved by the Technological Educational Institution of Athens research committee and followed the principles of the Helsinki Declaration and its later amendments.

Instruments

Western Ontario and McMaster LK 3.1 (WOMAC) index

In the present study, the Likert 3.1 Greek for Greece format of the WOMAC index was used [14]. The original Greek translation of the WOMAC LK3.1 Index was developed from the WOMAC LK3.1 English for Canada source Index using Standard Operating Procedures, under Professor Bellamy's copyright. The index is a 24-item questionnaire divided into three subscales, which measure pain (WOMAC-pain; 5 items, score range 0–20), stiffness (WOMAC-stiffness; 2 items, score range 0–8), and physical function (WOMAC-function; 17 items, score range 0–68). For the purposes of this study, each subscale item score was normalized on a 0–10 scale to correct for differences in scale length, so that a score of 0 represented the best health status and a score of 10 the worst. The three

normalized subscale values were summed to provide the normalized WOMAC-total score [14].

Medical outcomes study 36-item short-form health survey

The SF-36 is a general self-report instrument designed to be used in general and specific population groups for clinical practice and research. It is a 36-item questionnaire divided into eight subscales that measure eight different dimensions of health status, which represent the most frequently measured concepts in widely used health surveys and those most affected by disease and treatment [15]. The subscale item scores are coded, summed, and normalized to a scale ranging from 0 to 100. A score of 0 represents the worst health status and a score of 100 the best. In the present study, the physical functioning (SF36-function), role physical (SF36-role physical), and bodily pain (SF36-pain) subscales of the Greek SF-36® Health Survey (IQOLA SF-36 Greek Standard Version 1.0) were used [16, 17].

Visual analog scale/faces pain scale-revised (VAS/FPS-R)

The visual analog scale (VAS) is an instrument frequently used to measure pain intensity based on a 0–100 mm VAS [18]. In the present study, a combination of the Greek versions of VAS [19] and the faces pain scale-revised (FPS-R) [20, 21] scales was used. The FPS-R includes six facial expressions, which cover the entire range of pain levels in a hierarchical order and has a high degree of concurrent validity and a high correlation with VAS (0.829) [22]. The FPS-R was used in combination with the VAS, in preference to the VAS alone, because the instrument asks patients to describe their pain according to a facial expression that corresponds to their pain and enables them to translate their subjective experience of pain into a quantitative, numeric measure such as VAS.

Timed up and go test

The “timed up and go” (TUG) test was introduced in 1991 by Podsiadlo and Richardson [23] as a modification of the “get Up and go test” of Mathias et al. [24]. It is an easy to administer physical performance measure that requires minimal equipment and interpretation and has been widely used to describe and monitor functional mobility [25]. It assesses common problems found in people with lower extremity OA, incorporating four different subcomponents that represent different functioning constructs (walking, turning, rising from a chair, sitting down into a chair). The time it takes for a person to complete the test is correlated strongly with the level of his/her functional mobility [23]. TUG has been used in a number of studies for the functional evaluation of OA patients and appears to be a

responsive and useful outcome measure to guide clinical care for knee OA patients [7–9, 26]. During the test, all participants were asked to stand up from a chair without armrests and with a seat height of 44 cm, walk at a comfortable pace to a line 3 m away, cross the line and turn through 180°, walk back, and sit in their starting position. The use of an assistive device was allowed for those using one, but no verbal encouragement or personal assistance was given. After one pilot test, the average time of the two successive trials was recorded using a timer with an accuracy of 1/100 s.

Procedures

At initial assessment (day 1), all candidates completed a standardized questionnaire recording socio-demographic and personal data, together with information about their osteoarthritis-related history. Subsequently, the Greek version of WOMAC, and the SF36-function, SF36-role, and SF36-pain subscales of the Greek SF-36 Health Survey were given to all participants. Patients were then asked to draw a vertical line on the combined 100 mm visual analog scale/faces pain scale-revised (VAS/FPS-R) at a point that corresponded to their current level of pain at rest. Participants’ functional mobility was evaluated by their TUG test performance. Seven days after the initial assessment (day 8), WOMAC was re-administered to all participants so that the reliability properties of the instrument could be evaluated. Subject guidance and questionnaire completion were carried out under the supervision of the same member of the research team.

Data analysis

Statistical analyses were performed using the IBM® SPSS® version 19 software package (2010 SPSS Inc., an IBM Company, Chicago, IL, USA). The distribution and normality of the collected data were tested using the Kolmogorov–Smirnov test and probability–probability plots. Age, somatometrics, WOMAC and SF-36 outcomes, TUG, and VAS/FPS-R scores were all normally distributed and are presented as mean values \pm standard deviation. There were no missing data for any of the variables analyzed. In all analyses, significance was set at $p < 0.05$.

Reliability study

Internal consistency was assessed by Cronbach’s Alpha statistic. Test–retest reliability was assessed by computing intra-class correlation coefficients (ICC; two-way mixed–single measures, 2-1-1 model) with 95 % confidence intervals, between the day 1 and day 8 WOMAC outcomes. The WOMAC scores of the two assessments were also tested

for systematic differences (repeatability) using the paired samples *t* test and ANOVA. In all reliability testing, a threshold value of 0.70 was chosen.

Validity study

This study aimed to examine the construct (convergent, nomological and known-groups) and criterion-related (concurrent) validity of the WOMAC index in knee OA patients. In all validity analyses, the coefficient values were characterized as follows: 0.00–0.19 = poor, if any; 0.20–0.39 = fair; 0.40–0.59 = moderate; 0.60–0.79 = good; 0.80–1.00 = high/strong [27].

Construct validity The item-total correlations within each WOMAC subscale were compared in order to test whether all items of each subscale were related to the same construct. Convergent validity was evaluated by examining the correlations with other measures with similar constructs. Nomological validity was evaluated by calculating the inter-scale correlations to examine whether WOMAC subscales were distinct but related constructs. The known-groups validity was tested by independent samples *t* test to examine the ability of WOMAC-function to discriminate OA patients into two sub-groups based on their functional level at initial assessment. The TUG test was used as an external criterion, and the participants' functional status was determined by the completion time of this physical performance measure.

Criterion-related validity Pearson's correlations were used to test WOMAC for concurrent validity against TUG, SF-36, and VAS/FPS-R scores. In addition, partial correlations of the index outcomes against all validation criteria, adjusted for confounders, were derived. Multiple linear regression analyses were implemented to further investigate the index's predictive validity, by examining the associations between WOMAC outcomes (independent variables) and TUG, VAS/FPS-R, and SF-36 scores (dependent variables), taking account of possible confounders. Preliminary testing was done to check for violations of assumptions (normality, linearity, and homoscedasticity) and outliers. Initially, the assumption of normality for the dependent (validation criteria) and for the independent variables (WOMAC outcomes) was verified.

Results

Descriptives

Twenty-three individuals of the 150 randomly selected were excluded on the basis of the exclusion criteria and four did not turn up on the first evaluation day (day 1). Thus,

123 OA patients (67 women) with a mean age of 69.5 years participated and completed all the assessment protocols (Table 1). The distributions of the non-participants' demographic and personal characteristics were similar to those participating in the study. Descriptive data for WOMAC and SF-36 outcomes as well as for TUG and VAS/FPS-R scores are presented in Table 2. No statistically significant differences were found between the WOMAC scores at initial assessment and on re-assessment days. Finally, no alteration in the participants' clinical status was noted and no treatment interventions were delivered between assessments. All WOMAC scores were better in women on both assessment days, while all SF-36 outcomes were better in men; however, these differences were not statistically significant. In addition, no significant sex-related differences were found in the TUG and VAS/FPS-R scores (Table 2).

Reliability study

Internal consistency The Cronbach's Alpha of the Greek version of the WOMAC index was high (for pain and stiffness subscales) to excellent (for function subscale). The respective data are presented in online resource (Supplementary Table 1). All "if item deleted" values were lower than the respective subscale's overall Cronbach's Alpha, indicating that all items had to be included in the respective subscale and suggesting that all the index's items were interdependent and homogeneous in terms of the construct they measure. **Test-retest reliability:** Intra-class correlation coefficients (ICC) for test-retest reliability between day 1 and day 8 outcomes were excellent and ranged from 0.91 to 0.95 (entire group), and from 0.89 to 0.96 when the results were stratified by sex (Table 3). **Repeatability:** When paired samples *t* tests and ANOVA were applied, no systematic differences were found between the day 1 and day 8 WOMAC outcomes (Table 2).

Validity study

Construct validity

Acceptable validity was indicated by high to excellent (0.70–1.00) for 3 out of 5 item pairings of WOMAC-pain, 15 of the 17 of WOMAC-function, and both item pairings of WOMAC-stiffness (Table 4). **Convergent validity:** WOMAC-function showed a higher correlation with SF36-function than with SF36-pain; WOMAC-function showed a higher correlation with TUG than did WOMAC-pain; and WOMAC-pain showed a higher correlation with VAS/FPS-R than did WOMAC-function (Table 5). **Nomological validity:** The inter-scale correlations indicated that WOMAC subscales were moderately related (data are not shown). Since these relations did not exceed 0.80, there was no cause

Table 1 Characteristics of the study population

	All (n = 123)	Males (n = 56)	Females (n = 67)	p**
Personal data				
Age (year)*	69.5 ± 6.2	70.9 ± 5.6	68.2 ± 6.5	0.02
Height (cm)*	166.9 ± 9.4	172.3 ± 8.5	160.3 ± 5.9	<0.001
Weight (kg)*	79.2 ± 13.2	85.4 ± 13.9	73.8 ± 10.3	<0.001
BMI (kg/m ²)*	28.7 ± 3.9	28.7 ± 3.6	28.8 ± 4.2	NS
Dominant leg (%)				
Right	86.7	80.4	92.2	0.05
Left	13.3	19.6	7.8	0.05
Affected leg (%)				
Right	55.0	53.6	56.3	NS
Left	45.0	46.4	43.7	NS
Use of a cane (%)				
Frequency of use (%) rare/sometimes/often	0.8/2.5/14.2	1.8/3.6/17.9	1.6/10.9/12.5	NS

BMI body mass index

* The results are presented as mean ± standard deviation

** Differences in personal data between sexes

Table 2 WOMAC normalized outcomes on day 1 and day 8 assessments

	All (n = 123)	Males (n = 56)	Females (n = 67)	p*
Day 1				
WOMAC-total	7.9 ± 5.9	6.9 ± 5.3	8.8 ± 6.4	NS
WOMAC-pain	3.1 ± 2.0	2.8 ± 1.8	3.3 ± 2.1	NS
WOMAC-stiffness	1.9 ± 2.4	1.5 ± 2.0	2.2 ± 2.7	NS
WOMAC-function	3.0 ± 2.2	2.7 ± 2.1	3.3 ± 2.3	NS
TUG score	11.5 ± 2.2	11.4 ± 2.4	11.5 ± 2.1	NS
VAS/FPS-R score	3.4 ± 2.3	3.1 ± 2.1	3.6 ± 2.4	NS
SF36-function	57.9 ± 26.2	59.7 ± 27.3	56.3 ± 25.2	NS
SF36-role	53.1 ± 45.6	56.7 ± 46.1	50.0 ± 45.2	NS
SF36-pain	57.1 ± 25.1	60.0 ± 23.1	54.7 ± 26.7	NS
Day 8				
WOMAC-total	8.2 ± 5.4**	7.3 ± 4.8	8.9 ± 5.9	NS
WOMAC-pain	3.1 ± 1.9**	2.8 ± 1.8	3.3 ± 1.9	NS
WOMAC-stiffness	1.8 ± 2.3**	1.5 ± 2.1	2.2 ± 2.4	NS
WOMAC-function	3.2 ± 1.9**	2.9 ± 1.8	3.4 ± 2.0	NS

SF-36, TUG, and VAS/FPS-R scores on day 1 assessment

WOMAC WOMAC LK 3.1 Greek for Greece Index, SF-36 Short-Form 36 Health Survey, TUG timed up and go test, VAS/FPS-R combined visual analog/faces pain scale-revised

* Non-significant differences in WOMAC outcomes between sexes

** Non-significant differences in WOMAC outcomes between day 1 and day 8 assessments

for concern about multicollinearity. *Known-groups validity*: There is no agreement in the literature on a predefined numerical TUG value for the functional dichotomization of knee OA patients. For this reason, widely accepted statistical methods aiming to determine a meaningful cutoff value for TUG were implemented in the present study. Subsequently, the mean completion time (11.49 s) of our study population

(meaningful cutoff value) was used to dichotomize the participants into two functional groups (good functional status, TUG < 11.49 s vs. poor functional status, TUG ≥ 11.49 s). Analysis of the data showed that there were significant differences between the mean WOMAC-function scores in the two functional groups ($t = -7.734, p < 0.001$), indicating that the subscale was able to dichotomize knee OA patients according to their functional status.

Criterion-related validity

Concurrent validity Pearson’s correlations between the parametric WOMAC outcomes against validation criteria are presented in online resource (Supplementary Table 2). The possible involvement of personal characteristics as covariates in the association between WOMAC outcomes and validation criteria was investigated. Multiple linear regression analyses indicated that “age” and “use of an assistive device” were the most significant confounders in the association between WOMAC outcomes and validation criteria. Accordingly, partial correlation coefficients, adjusted for confounders, were calculated (Table 5). In the overall population, WOMAC-total was significantly associated with all validation criteria. The inverse direction of the correlations between WOMAC and SF-36 outcomes can be attributed to the fact that higher values in WOMAC indicate a worse health status, whereas higher values in SF-36 indicate better health. Good to high correlations were found between WOMAC and SF36-function and SF36-pain, while fair to moderate correlations were found between WOMAC and SF36-role physical. WOMAC-function was more strongly correlated with SF36-function ($r = -0.86, p < 0.001$), TUG ($r = 0.71, p < 0.001$), and SF36-pain ($r = -0.71, p < 0.001$). WOMAC-pain was more strongly correlated with SF36-function ($r = -0.72, p < 0.001$),

Table 3 WOMAC test–retest reliability correlations between day 1 and day 8 assessments

	All (<i>n</i> = 123)	Cronbach's α^*	Males (<i>n</i> = 56)	Females (<i>n</i> = 67)
	ICC (95 % CI)		ICC (95 % CI)	ICC (95 % CI)
WOMAC-total	0.95 (0.93–0.97)	0.98	0.94 (0.90–0.96)	0.96 (0.93–0.98)
WOMAC-pain	0.92 (0.89–0.95)	0.96	0.91 (0.85–0.95)	0.93 (0.89–0.96)
WOMAC-stiffness	0.91 (0.87–0.93)	0.95	0.89 (0.82–0.93)	0.91 (0.86–0.95)
WOMAC-function	0.93 (0.89–0.96)	0.97	0.91 (0.82–0.95)	0.95 (0.91–0.97)

All correlations were significant at $p < 0.001$

WOMAC WOMAC LK 3.1 Greek for Greece Index, ICC intra-class correlation coefficients, CI confidence intervals

* Cronbach's α coefficients for all subjects

Table 4 Item-total correlations within each subscale of the Greek version of the WOMAC index

	Items	Pearson correlation
WOMAC-pain	1	0.841
	2	0.796
	3	0.641
	4	0.639
	5	0.821
WOMAC-stiffness	1	0.942
	2	0.897
WOMAC-function	1	0.723
	2	0.729
	3	0.750
	4	0.745
	5	0.680
	6	0.802
	7	0.762
	8	0.833
	9	0.840
	10	0.765
	11	0.833
	12	0.624
	13	0.837
	14	0.713
	15	0.817
	16	0.777
	17	0.822

All correlations were significant at the 0.01 level (2 tailed)

WOMAC WOMAC LK 3.1 Greek for Greece Index

VAS/FPS-R ($r = 0.71$, $p < 0.001$), and SF36-pain ($r = -0.67$, $p < 0.001$). Of all WOMAC outcomes, stiffness subscale had the lowest, though still significant, correlation with all validation criteria in all study subgroups examined. When data were stratified by sex, WOMAC outcomes were also significantly associated with all validation criteria, with the correlations ranging from moderate to excellent

Table 5 Partial correlations between WOMAC outcomes and SF-36, TUG, and VAS/FPS-R scores on day 1, adjusted for age, sex, and use of an assistive device

	TUG	VAS/ FPS-R	SF36- function	SF36-role	SF36- pain
Day 1					
WOMAC-total	0.632*	0.631*	-0.805*	-0.389*	-0.685*
WOMAC-pain	0.566*	0.705*	-0.715*	-0.346**	-0.668*
WOMAC- stiffness	0.431*	0.413*	-0.596*	-0.327**	-0.477*
WOMAC- function	0.712*	0.609*	-0.860*	-0.367*	-0.713*
TUG	–	0.619*	-0.661*	-0.251***	-0.435*
VAS/FPS-R	0.619*	–	-0.588*	-0.425*	-0.541*

WOMAC WOMAC LK 3.1 Greek for Greece Index, SF-36 Short-Form 36 Health Survey, TUG timed up and go test, VAS/FPS-R combined visual analog/faces pain scale-revised

* $p \leq 0.001$; ** $p < 0.01$; *** $p < 0.05$

(data not shown). In multiple linear regression analyses, all personal characteristics, together with the WOMAC subscales, were entered into the model. WOMAC-total was excluded because of multicollinearity problems. WOMAC-function was a significant factor for TUG, WOMAC-pain for VAS/FPS-R, and both for SF36-function and SF36-pain (Table 6). All metric variables included in the analysis satisfied the assumption of normality. Stepwise analyses (bidirectional elimination) were applied to provide the best regression models for all validation criteria. In the resulting models, age and the use of an assistive device were entered as covariates for TUG, VAS/FPS-R, and SF-36 scores.

Discussion

This is among the very few studies [7, 9], and the first in Greece, where both self-reported and physical performance measures were used to examine the clinimetric properties of the WOMAC index in OA patients. The index was found

Table 6 Multiple linear regression analysis of WOMAC outcomes with validation criteria

	Beta coefficient (standard error)	Standardized beta	<i>p</i>
TUG score			
WOMAC-pain	0.077 (0.133)	0.069	0.564
WOMAC-stiffness	−0.079 (0.086)	−0.088	0.358
WOMAC-function	0.580 (0.134)	0.586	<0.001
VAS/FPS-R			
WOMAC-pain	0.683 (0.154)	0.600	<0.001
WOMAC-stiffness	−0.057 (0.093)	−0.066	0.544
WOMAC-function	0.124 (0.146)	0.132	0.398
SF36-function			
WOMAC-pain	−1.979 (1.032)	−0.149	<0.05
WOMAC-stiffness	−0.212 (0.667)	−0.020	0.751
WOMAC-function	−8.166 (1.043)	−0.692	<0.001
SF36-pain			
WOMAC-pain	−3.296 (1.400)	−0.259	<0.05
WOMAC-stiffness	−0.044 (0.905)	−0.004	0.961
WOMAC-function	−5.878 (1.415)	−0.519	<0.001

WOMAC WOMAC LK 3.1 Greek for Greece Index, SF-36 Short-Form 36 Health Survey, TUG timed up and go test, VAS/FPS-R combined visual analog/faces pain scale-revised

to have excellent reliability properties and presented significant validity against the TUG test, VAS/FPS-R, and the SF-36 subscales used.

Reliability study

Our results indicated that the Greek version of WOMAC was remarkably consistent between the two measurements. The instrument was assessed in terms of internal consistency and test–retest reliability. Both overall Cronbach's α coefficient and the ICC values were found to be excellent, similar to those reported for other versions of WOMAC [28–32], indicating that the index has high reliability properties in patients with knee OA. The item-total and item-subscale correlations were high to excellent, showing that the instrument can be reliably used to measure all three parameters it was designed for: pain, stiffness, and physical function. A higher correlation value was derived, during WOMAC-pain item analysis, when item 3 (nocturnal pain) was omitted, possibly due to the variability of the pain intensity in knee OA [33].

Validity study

Construct validity

The construct validity of WOMAC was established by the findings of convergent, nomological, and known-groups

validity analyses. The high to excellent item-total correlations within each subscale confirmed the construct validity of the index, providing evidence that the items of each subscale were strongly related to the same construct. In addition, the subscales showed the expected pattern of correlations with the set of the external measures of the same construct used, upholding the instrument's *convergent validity*. WOMAC subscales were found to assess related constructs, but they were sufficiently unique not to be considered redundant, confirming the *nomological validity* of the instrument. *Known-group analysis* of the data showed that WOMAC-function can be used to distinguish OA patients according to their functional status, based on the TUG test performance time.

Criterion-related validity

Concurrent validity WOMAC outcomes were significantly associated with all validation criteria, presenting moderate to high correlation coefficients, confirming the instrument's concurrent validity. The weaker association between WOMAC-stiffness and comparator measures (WOMAC-stiffness shows generally lower levels of correlation) [14] is attributable to the nature of stiffness, which is easily discernible by patients with knee OA and is distinct from both pain and disability. Since none of the comparator measures in this study contained elements specific to the evaluation of perceived joint stiffness, there was necessarily a weaker association between WOMAC-stiffness and these other measures. The lower validity coefficients found between WOMAC outcomes and SF36-role physical are in line with those of other validity studies [28, 32, 34] and may be explained by the fact that WOMAC was designed as a measure of functional disability and pain, rather than a generic measure of health status that assesses physical role, among other factors. The combined VAS/FPS-R scale was used as an additional self-report validation criterion. Apart from VAS/FPS-R versus WOMAC-stiffness, our results showed good correlations, in line with the findings of Faucker et al. [35]. However, compared to the studies of Barasan et al. [32] and Guermazzi et al. [36], our findings exhibit higher values, possibly due to the use of the combined VAS/FPS-R, which enables participants to self-characterize the sense of pain in a more convenient way.

In our study, the pain domain of the WOMAC index was more strongly correlated with the SF36-function than with SF36-pain. These differences could possibly be due to the differences in the site specificity and recall period constructs between the pain subscales of the two instruments: (a) SF36-pain lacks any site specificity, while WOMAC-pain is joint specific; and (b) the very different recall periods for SF-36 (4 weeks) and the WOMAC index (48 h). These two factors together may, in some studies [29, 37]

but not in others [30, 32, 38], result in disordered patterns of correlation that are likely related to experiential factors (involving levels of pain and function over the longer term affecting the SF-36), and the varying extent of multijoint involvement among different patients. Therefore, the provided pattern of correlations upholds both the construct and the concurrent validity of the WOMAC Index, while the paradoxical observations regarding the WOMAC-pain versus SF36-function and SF36-pain correlations are explainable and do not detract from the instrument's validity.

Finally, multiple linear regression analyses indicated that WOMAC-function was a significant factor for TUG, WOMAC-pain for VAS/FPS-R, and both for SF36-function and SF36-pain, upholding the index's predictive validity. This finding is reinforced by the good to high inter-correlations presented in Table 5.

TUG test as external criterion

Our primary focus was to use the TUG test as a complementary objective criterion in order to test the validity of the physical function domain of the WOMAC index. In addition, since the dimensions of pain and stiffness subscales of the WOMAC index are closely related with the patients' functional mobility, we sought to explore the validity correlations between the total WOMAC index and the TUG test. As Wright et al. pointed out (2010), physical performance tests, including TUG, assess differing dimensions of physical function than does the self-reported WOMAC index (patient's ability vs. patient's perception). In addition, physical performance measures are good predictors of physical functioning [7] and compensate for the problems inherent in self-report measures, such as the ability and willingness to answer questions correctly [39]. The two assessment methods provide complementary information [9], and they are both needed to perceive the multidimensional impact of pain [6] and to capture the construct of physical function in its entirety [8], which are essential to clinical research and practice. The WOMAC–TUG correlations were found to be moderate to good. Moreover, both TUG versus WOMAC-function (0.71) and versus WOMAC-total (0.63) correlations are among the highest reported coefficients between WOMAC outcomes and other physical performance measures [7–9]. Thus, our findings indicated that the TUG test was a good choice to use as an additional validation criterion, implying its valuable contribution in capturing the construct of physical function of the knee OA patients in its entirety.

Strengths and limitations

The random selection of the participants from a well-defined cohort, with demographic characteristics similar to

those of the general population, was an important strength of this study. In addition, conducting an extensive reliability study and examining the validity properties of the instrument against both self-report and physical performance measures added statistical power to our results. However, the use of a battery of physical performance measures instead of one would have been added more strength in the present study. On the other hand, some potential limitations should be noted. The WOMAC index was validated only in specific joint and age groups and that may influence the extent to which our results can be generalized. A second limitation is that the sample for the present WOMAC validity testing was marginal, being close to the lowest acceptable population size, based on the "rule of five" [40] and on sampling estimates reported in other WOMAC validation studies [28–30, 33, 34]. Finally, an examination of the index's responsiveness was not included.

Conclusions

The WOMAC 3.1 index was found to be a reliable and valid assessment tool that can be used to evaluate patients with knee OA, showing excellent reliability and significant fair to strong validity properties. A broader awareness of these findings in the Greek setting would facilitate objective comparisons between studies of different national origin and would contribute to the validity of future meta-analyses.

Acknowledgments The authors would like to thank Philip Lees, Technical Editor of the Hellenic Journal of Cardiology, for his invaluable editorial assistant with the English text.

Conflict of interest Nicholas Bellamy is the registered copyright and registered trademark holder of the WOMAC® Index. None of the other authors had any conflict of interest in relation to this study.

References

1. Arthritis Foundation® (2008) Osteoarthritis fact sheet. http://www.arthritis.org/files/images/newsroom/media-kits/Osteoarthritis_fact_sheet.pdf
2. Zhang Y, Jordan JM (2010) Epidemiology of osteoarthritis. *Clin Geriatr Med* 26(3):355–369
3. Bellamy N, Buchanan WW (1986) A preliminary evaluation of the dimensionality and clinical importance of pain and disability in osteoarthritis of the hip and knee. *Clin Rheumatol* 5(2):231–241
4. Bellamy N, Buchanan W, Goldsmith C, Campbell J, Stitt L (1988) 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* 15(12):1833–1840
5. Bellamy N (1997) Osteoarthritis clinical trials: candidate variables and clinimetric properties. *J Rheumatol* 24:768–778

6. Stratford PW, Kennedy DM, Woodhouse LJ (2006) Performance measures provide assessment of pain and function in people with advanced osteoarthritis of the hip or knee. *Phys Ther* 86:1489–1496
7. Juhakoski R, Tenhonen S, Anttonen T, Kauppinen T, Arokoski JP (2008) Factors affecting self-reported pain and physical function in patients with hip osteoarthritis. *Arch Phys Med Rehabil* 89:1066–1073
8. Wright AA, Cook CE, Baxter GD, Garcia J, Abbott JH (2010) Relationship between the Western Ontario and McMaster Universities Osteoarthritis Index Physical Function Subscale and physical performance measures in patients with hip osteoarthritis. *Arch Phys Med Rehabil* 91:1558–1564
9. Stratford PW, Kennedy DM (2006) Performance measures were necessary to obtain a complete picture of osteoarthritic patients. *J Clin Epidemiol* 59:160–167
10. Andrianakos AA, Kontelis LK, Karamitsos DG, Aslanidis SI, Georgountzos AI, Kaziolas GO et al (2006) Prevalence of symptomatic knee, hand, and hip osteoarthritis in Greece. The ESOR-DIG study. *J Rheumatol* 33:2507–2514
11. Altman R, Asch E, Bloch D, Bole G, Borenstein D, Brandt K et al (1986) Development of criteria for the classification and reporting of osteoarthritis: classification of osteoarthritis of the knee. *Arthritis Rheum* 29:1039–1049
12. Puopolo A, Boice J, Fidelholtz J, Littlejohn T, Miranda P, Berrocal A et al (2007) A randomized placebo-controlled trial comparing the efficacy of etoricoxib 30 mg and ibuprofen 2400 mg for the treatment of patients with osteoarthritis. *Osteoarthr Cartil* 15(12):1348–1356
13. Wood L, Peat G, Mullis R, Thomas E, Foster N (2011) Impairment-targeted exercises for older adults with knee pain: protocol for a proof-of-principle study. *BMC Musculoskelet Disord* 12:2
14. Bellamy N (2011) WOMAC® osteoarthritis index: user guide X. Prof. Nicholas Bellamy, Herston, Australia
15. Ware J, Sherbourne C (1992) The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 30(6):473–483
16. Pappa E, Kontodimopoulos N, Niakas D (2005) Validating and norming of the Greek SF-36 Health Survey. *Qual Life Res* 14(5):1433–1438
17. Anagnostopoulos F, Niakas D, Pappa E (2005) Construct validation of the Greek SF-36 health survey. *Qual Life Res* 14(8):1959–1965
18. Flandry F, Hunt J, Terry G, Hughston J (1991) Analysis of subjective knee complaints using visual analogue scales. *Am J Sports Med* 19(2):112–118
19. Mpouzika MDA, Kletsiou E, Petsa C, Papatthanassoglou EDE (2009) Validation of the Greek version of two behavioural pain scales in critically ill patients. *Nosileftiki* 48(3):308–316
20. Hicks CL, von Baeyer CL, Spafford P, van Korlaar I, Goode-nough B (2001) The Faces Pain Scale-Revised: toward a common metric in pediatric pain measurement. *Pain* 93:173–183
21. von Baeyer C, Wood C, Jaaniste T (2009) Instructions for administering the Faces Pain Scale-Revised (FPS-R) in languages other than English. Edition 6. <http://www.sfap.org/pdf/VIII-19e-pdf.pdf>
22. Stuppy D (1998) The Faces Pain Scale: reliability and validity with mature adults. *Appl Nurs Res* 11(2):84–89
23. Podsiadlo D, Richardson S (1991) The timed “Up & Go”: a test of basic functional mobility for frail elderly persons. *J Am Geriatr Soc* 39:142–148
24. Mathias S, Nayak U, Isaacs B (1986) Balance in the elderly patients: the “get-up and go” test. *Arch Phys Med Rehabil* 67(6):387–389
25. VanSwearingen J, Brach J (2001) Making geriatric assessment work: selecting useful measures. *Phys Ther* 81:1233–1252
26. Cecchi F, Molino-Lova R, Di Iorio A, Conti AA, Mannoni A, Lauretani F et al (2009) Measures of physical performance capture the excess disability associated with hip pain or knee pain in older persons. *J Gerontol A Biol Sci Med Sci* 64:1316–1324
27. Fowler J, Jarvis P, Chevannes M (2002) *Practical statistics for nursing and health care*. Wiley, West Sussex
28. Stucki G, Sangha O, Stucki S, Michel B, Tyndall A, Dick W et al (1998) Comparison of the WOMAC (Western Ontario and McMaster Universities) osteoarthritis index and a self-report format of the self-administered Lequesne–Algofunctional index in patients with knee or hip osteoarthritis. *Osteoarthr Cartil* 6(2):79–86
29. Salaffi F, Learardini G, Canesi B, Mannoni A, Fioravanti A, Caporali R et al (2003) Reliability and validity of the Western Ontario and McMaster Universities (WOMAC) Osteoarthritis Index in Italian patients with osteoarthritis of the knee. *Osteoarthr Cartil* 11(8):551–560
30. Kuptniratsaikul V, Rattanachaiyanont M (2007) Validation of a modified Thai version of the Western Ontario and MacMaster (WOMAC) osteoarthritis index for knee osteoarthritis. *Clin Rheumatol* 26(10):1641–1645
31. Faik A, Benbouazza K, Amine B, Maaroufi H, Bahiri R, Lazrak N et al (2008) Translation and validation of Moroccan Western Ontario and McMaster Universities (WOMAC) osteoarthritis index in knee osteoarthritis. *Rheumatol Int* 28(7):677–683
32. Basaran S, Guzel R, Seydaoglu G, Gyler-Uysal F (2010) Validity, reliability, and comparison of the WOMAC osteoarthritis index and Lequesne algofunctional index in Turkish patients with hip or knee osteoarthritis. *Clin Reumatol* 29(7):749–756
33. Woolhead G, Goberman-Hill R, Dieppe P, Hawker G (2010) Night pain in hip and knee osteoarthritis: a focus group study. *Arthritis Care Res (Hoboken)* 62(7):944–949
34. Escobar A, Quintana J, Bilbao A, Azkarate J, Guenaga J (2002) Validation of the Spanish version of the WOMAC questionnaire for patients with hip or knee osteoarthritis. *Clin Rheumatol* 21(6):466–471
35. Faucher M, Poiraudou S, Lefevre-Colau M, Rannou F, Fermanian J, Revel M (2004) Assessment of the test–retest reliability and construct validity of a modified WOMAC index in knee osteoarthritis. *Joint Bone Spine* 71(2):121–127
36. Guermazi M, Poiraudou S, Yahia M, Mezganni M, Fermanian J, Habib Elleuch M et al (2004) Translation, adaption and validation of the Western Ontario and McMaster Universities osteoarthritis index (WOMAC) for an Arab population: the Sfax modified WOMAC. *Osteoarthr Cartil* 12(6):459–468
37. Xie F, Li SC, Goeree R, Tarride JE, O’Reilly D, Lo NN et al (2008) Validation of Chinese Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) in patients scheduled for total knee replacement. *Qual Life Res* 17:595–601
38. Thumboo J, Chew LH, Soh CH (2001) Validation of the Western Ontario and McMaster University osteoarthritis index in Asians with osteoarthritis in Singapore. *Osteoarthr Cartil* 9:440–446
39. Rockwood K, Awalt E, Carver D, MacKnight C (2000) Feasibility and measurement properties of the functional reach and the timed up and go tests in the Canadian study of health and aging. *J Gerontol A Biol Sci Med Sci* 55:M70–M73
40. Gorsuch RL (1983) *Factor analysis*, 2nd edn. L Erlbaum, Hillsdale, NJ