OsteoArthritis and Cartilage (2003) **11,** 551–560

doi:10.1016/S1063-4584(03)00089-X

Osteoarthritis and Cartilage

International Cartilage Repair Society

OARSI OSTEOARTHRITIS RESEARCH SOCIETY INTERNATIONAL

Reliability and validity of the Western Ontario and McMaster Universities (WOMAC) Osteoarthritis Index in Italian patients with osteoarthritis of the knee

F. Salaffi†*, G. Leardini‡, B. Canesi§, A. Mannoni‖, A. Fioravanti¶, R. Caporali††, G. Lapadula‡‡ and L. Punzi§§, on behalf of GOnarthrosis and Quality Of Life Assessment (GOQOLA) Study Group

†*Clinica Reumatologica, Università di Ancona, Azienda Sanitaria di Jesi, Jesi (Ancona), Italy*
‡*Divisione di Reumatologia, Venezia, Italy*
§*Divisione di Reumatologia, Azienda Ospedaliera Niguarda Ca' Granda, Milano, Italy*
‖*Unità Operativa di Reumatologia, Azienda Sanitaria di Firenze, Firenze, Italy*
¶*Istituto di Reumatologia, Università di Siena, Siena, Italy*
††*Cattedra e Unità Operativa di Reumatologia, Università di Pavia, IRCCS Policlinico San Matteo, Pavia, Italy*
‡‡*Cattedra e Divisione di Reumatologia, Università di Bari, Azienda Ospedaliera Policlinico di Bari, Bari, Italy*
§§*Cattedra e Divisione di Reumatologia, Università di Padova, Padova, Italy*

Summary

Objective: The Western Ontario and McMaster Universities (WOMAC) Osteoarthritis (OA) Index is a tested questionnaire to assess symptoms and physical functional disability in patients with OA of the knee and the hip. We adapted the WOMAC for the Italian language and tested its metric properties in 304 patients with symptomatic OA of the knee.

Methods: Three hundred and four consecutive patients, attending 29 rheumatologic outpatient clinic in northern, central, and southern Italy, were asked to answer two disease-specific questionnaires (WOMAC and Lequesne algofunctional index) and one generic instrument (Medical Outcomes Study SF-36 Health Survey—MOS SF-36). A sample of 258 patients was readministered the WOMAC 7–10 days after the first visit and the structured interview, which also assessed demographic and other characteristics. Internal consistency was assessed using Cronbach's alpha, reliability using intraclass correlation coefficients (ICCs), and construct and discriminant validity using Spearman's correlations, Wilcoxon rank sum test, and Kruskal–Wallis test.

Results: All WOMAC subscales (pain, stiffness, and physical function) were internally consistent with Cronbach's coefficient alpha of 0.91, 0.81, and 0.84, respectively. Test–retest reliability was satisfactory with ICCs of 0.86, 0.68, and 0.89, respectively. In comparison with the SF-36, the expected correlations were found when comparing items measuring similar constructs, supporting the concepts of convergent construct validity. Very high correlations were also obtained between WOMAC scores and Lequesne OA algofunctional index. WOMAC physical function, but not WOMAC stiffness and pain subscales, was weakly associated with radiological OA severity ($P=0.03$). Also, WOMAC pain score was inversely correlated ($P=0.01$) with years of formal education. Examination of discriminant validity showed that the scores on the WOMAC and SF-36 followed hypothesized patterns: the WOMAC discriminated better among subjects with varying severity of knee problems, whereas the SF-36 discriminated better among subjects with varying levels of self-reported health status and comorbidity.

Conclusion: The Italian version of WOMAC is a reliable and valid instrument for evaluating the severity of OA of the knee, with metric properties in agreement with the original, widely used version.

Key words: Osteoarthritis of the knee, Western Ontario and McMaster Universities (WOMAC) Osteoarthritis Index, Health status, Trial methodology.

Introduction

Osteoarthritis (OA) is a major cause of musculoskeletal pain, the single relevant cause of disability and handicap from arthritis, and an important community healthcare burden, in lost time at work and early retirement[1–4]. The knee joint is a common site of OA[5,6], and subjects with knee OA exhibit a characteristic pattern of decrements in function, generally concerning mobility, transfer from seated or supine position to standing, and activities of daily living (ADLs) involving the lower extremities[1,7]. The clinical metrology of OA is complex because, like other conditions, such as rheumatoid arthritis, ankylosing spondylitis, and fibromyalgia, there are few constants in the clinical presentation. Furthermore, OA may be symptomatic or not, and

This study was supported by Rottapharm s.r.l. and Pharmacia–Searle, Italy.
*Address correspondence and reprint requests to: Fausto Salaffi, Clinica Reumatologica, Università di Ancona, Ospedale Augusto Murri, Via dei Colli, 52, 60035 Jesi (Ancona), Italy. Tel: 39-731-534128 Ext. 32; Fax: 39-731-534124; E-mail: fsalaff@tin.it
Received 30 September 2002; revision accepted 15 April 2003.

the associated radiographs either normal or abnormal. In this context, it is necessary to identify valid and acceptable outcome measures in order to correctly evaluate the effectiveness of the therapy in OA. Such measures should benefit not only clinicians managing OA and purchasers of health care for this condition, but also, ultimately, the patients through improved forms of treatment. It is increasingly recognized that a key outcome measure for any health-care intervention for OA, as for many other conditions, is a change in health-related quality of life (HR-QoL)[8,9]. Quality of life has multiple dimensions, and in recent years, researchers and clinicians have produced several generic and disease-specific questionnaires to measure it[10–12]. Generic measures that have been extensively tested and shown to be reliable and valid in different populations and for a variety of uses include the SF-36[13], the Sickness Impact Profile[14], the Nottingham Health Profile[15], and the EuroQuol[16]. The most widely used condition-specific instruments for the assessment of hip or knee OA is the Western Ontario and McMaster Universities (WOMAC) OA Index[17,18], which is recommended by the Outcome Measures in Rheumatoid Arthritis Clinical Trials (OMERACT)[19,20]. The WOMAC scale was designed to measure dysfunction and pain associated with OA of the lower extremities by assessing 17 functional activities, five pain-related activities, and two stiffness categories[21]. This self-assessment multidimensional instrument has been well studied, and many of its psychometric properties are known[19]. However, to enable comparison between assessments made in different countries, this questionnaire needs not only to be translated, but also to be adapted for use in different cultures. We report on the linguistic validation of an Italian version of WOMAC OA Index, and present data on its metric properties.

Methods

PATIENTS

Recruitment of patients

In this cross-sectional study, the WOMAC was administered to 304 outpatients with symptomatic tibiofemoral OA of the knee, fulfilling the American College of Rheumatology (ACR) criteria for knee OA[22], enrolled in 29 rheumatologic centers in northern, central, and southern Italy. The multicentric feature of this validation study is an important issue in that the centers were chosen to minimize any possible bias due to different cultural, semantic, and demographic factors. During the clinic visit, each patient underwent a complete assessment according to a standard protocol. A sample of 258 patients was readministered the WOMAC 7–10 days after the first visit to evaluate the instrument's test–retest reliability. We based classification of knee radiographs on the standard Kellgren/Lawrence (K/L) criteria[23] (graded 0–4, where 0, absence of any sign of radiological OA (ROA); 1, possible osteophytes only; 2, definite osteophytes and possible joint space narrowing; 3, moderate osteophytes and/or definite joint space narrowing; and 4, large osteophytes, severe joint space narrowing, and/or bony sclerosis). ROA is defined as a score higher than 2, and severe ROA as a score higher than 3 in the left and/or right joint. Radiographs used in this study were generally obtained within 1 year of the date of the questionnaire assessments[24]. To be eligible, all patients had to be symptomatic, requiring either nonsteroidal anti-inflammatory drugs (NSAID) or a pure analgesic, or both to control their pain. The knee designated as the 'study joint' was the primary source of pain or disability in the lower extremity. Exclusion criteria were as follows: concurrent systemic inflammatory rheumatic disease; medical comorbidity that would render the patient unable to participate fully in study procedures (e.g., terminal conditions, such as end-stage renal disease, heart failure, or malignancy); alcohol abuse or a psychiatric disorder; and previous or planned knee arthroplasty of the study joint. All centers had approval from their respective ethics committees. All patients provided informed consent.

Background and illness-related variables

Demographic and socio-economic information were assessed from the interview with patient. Age is given in years. Educational level was separated into three categories based on the Italian school system: 1, primary school; 2, secondary school; and 3, high school or university. Marital status was recorded in two categories: 1, living together and 0, living alone. The body mass index (BMI; body weight divided by the square of the height) was used to assess overweight, which is a known risk factor for OA of the knee. Being overweight was defined as having a BMI of 26–29 kg/m^2 and being obese having a BMI >30 kg/m^2. In all patients was assessed the presence of comorbidities. These were ascertained through patient's self-reports using additional questions probed for the presence of nine specific comorbid conditions (hypertension, myocardial infarction, lower extremity arterial disease, major neurologic problem, diabetes, gastrointestinal disease, chronic respiratory disease, kidney disease, and poor vision). The total comorbidity score was the sum of the comorbidity conditions (0, not present; 1, present). This score ranged from 0 to 9.

MEASURES

WOMAC

The WOMAC is a disease-specific self-report multidimensional questionnaire assessing pain, stiffness, and physical functional disability[17,18]. This index has gained growing acceptance in OA assessment since its introduction in 1986. The pain dimension or scale includes five items asking pain at activity or rest. The stiffness dimension includes two questions. The function dimension explores the degree of difficulty in 17 activities. The original WOMAC is available in two formats, visual analog scales (VAS) and five Likert boxes, with similar metric properties[17,18,21]. The translation of the WOMAC in Italian format (for a description of the formats see Appendix) was done by two bilingual researchers aware of the objective of the questionnaire. It was then translated back into English by two different bilingual persons, who had no prior knowledge of the instrument. No major cultural adaptations were made. In this study, the Italian WOMAC was used in its VAS format, and all 24 items are rated by the subject on a 100 mm VAS ranging from 0 (indicating no pain, stiffness, or difficulty) to 100 (indicating extreme pain, stiffness, or difficulty). The range of the WOMAC scores is: pain (0–500); stiffness (0–200), and function (0–1700).

SF-36

The SF-36 is a generic instrument with scores that are based on responses to individual questions, which are

summarized into eight scales, each of which measures a health concept[13]. These scales include function domains and aspects of well being as follows: physical function; role limitations due to physical problems; body pain; vitality or energy level; role limitations due to personal or emotional problems; mental health; social function; and general health perception[13]. The physical function scale consists of 10 items that ask about involvement in a range of activities, such as running, playing, lifting heavy objects, climbing stairs, walking, and bathing or dressing oneself. Respondents are asked to rate on a three-point scale the extent to which their health limited their ability to engage in the various activities over the past 4 weeks (1, limited a lot; 2, limited a little, and 3, not limited at all). The pain scale consists of two items asking patient to rate pain severity over the past week on a five-point scale. The psychological and social function scales include five and two items, respectively. For each of the SF-36 scales, necessary items are recorded so that higher values indicate better health, and are then summed. The summed scores are transformed to a 0–100 scale, following its designated scoring algorithm, with higher scores reflecting better quality of life. These eight scales, weighted according to normative data, are scored from 0 to 100, with higher scores reflecting better quality of life[13]. The SF-36 survey also includes a single-item measure of health transition, which is not used to score any of the eight multi-items scales. The SF-36 has been validated for use in Italy[25], and it can be completed within 10 min by most people. Recently, the originators of the SF-36 have developed algorithms to calculate two psychometrically based summary measures: the Physical Component Summary Scale Score (PCS) and the Mental Component Summary Scale Score (MCS)[26,27]. The PCS and MCS provide greater precision, reduce the number of statistical comparisons needed, and eliminate the floor and ceiling effects noted in several of the subscales[28–30]. For the analysis, body pain, physical functioning, and MCS were selected.

Lequesne algofunctional index

The index contain three components: pain or discomfort; maximum distance walked; and ADLs[31,32]. Points are allocated according to response so that higher values indicate greater severity. The theoretical maximum score is 24. Although the metric properties of the Lequesne algofunctional index have been established[31,32], separate subsections have not been validated for independent application[33]. The Lequesne OA index was proposed as an interview technique[33].

Statistical analysis

Patients were included in any analysis only when relevant data were complete, using two statistical packages (Statistica for Macintosh, StatSoft, Inc., USA, and MedCalc version 6 for Windows, MedCalc Software, Belgium). Parametric techniques may be applicable for certain ordinal level data; however, our data were generally not normally distributed (Kolmogorov–Smirnov test for normal distribution), and therefore, the use of nonparametric techniques provided a more conservative estimate of statistical significance. Where appropriate, median and interquartile ranges are presented as well as means and standard deviations (SD). We also calculated the percentage of the sample achieving the lowest (floor effect) and highest (ceiling effect) possible WOMAC scores.

Reliability

Controversy exists over the relative merits of test–retest and internal consistency methods of assessing reliability[34]. In this study, the test–retest reliability of the WOMAC was analyzed using intraclass correlation coefficients (ICCs). The ICC reflects both systematic and random differences in test scores[35]. Values of ICC thus vary from 1 (perfectly reliable) to 0 (totally unreliable). The ICC was chosen in preference to the Pearson correlation, which may overestimate reliability[34]. We also assessed reliability in terms of internal consistency of the WOMAC subscales (pain, stiffness, and function). Internal consistency measures the extent to which items within a scale are correlated with each other[35]. If the WOMAC is internally consistent in the OA population, we would expect items within the individual scales (or dimension) to be highly correlated with each other. The Cronbach alpha statistic[36] is used to estimate the average of the correlations between items within a dimension. According to Steiner and Norman[37], a value of 0.8 is usually regarded as acceptable.

Validity

Establishing the criterion or content validity of an instrument claiming to measure HRQoL is difficult, as there are no established gold standards for comparison. Evidence for construct validity can only be accumulated by *a priori* hypothesized patterns of associations with other validated instruments[38]. In this study, the construct validity was examined in terms of convergence between similar dimensions of the SF-36 questionnaire, Lequesne algofunctional index and global health status. Correlations were made using Spearman's rank method. To investigate a possible influence of patient characteristics, such as age, sex, marital status, level of education, BMI, and radiographic OA severity on the WOMAC, the associations between the WOMAC subscales and these characteristics were quantified by Wilcoxon rank sum test and by Kruskal–Wallis one-way analysis of variance. Discriminant validity was assessed by comparing WOMAC scores in patients with and without other health conditions. These were ascertained through patient's self-reports using nine additional questions, also included in the questionnaire. For each dimension, the Spearman's correlation coefficient was calculated to assess how well each dimension correlates with the number of comorbidities.

Results

COHORT DISTRIBUTION

Table I shows the main socio-demographic characteristics of patients (age, sex, disease duration, BMI, the number and percentage of patients with K/L rating score, the number of comorbid conditions, and educational level). The mean age of the 304 patients examined was 65.7±9.3 years (range 50–82). Of the total respondents, 214 were females (70%), 90 were males (30%). The mean duration of OA was 9.5±7.9 years. The school education level was generally low: 62.5% had received only a primary school education and only 16.5% had received a high school education. The majority of patients (84.5%) were married and lived with the family; up to 46.5% of the patients were housewives. BMI, indicative of overweight, was recorded for 68% of the patients examined. Of the 304 subjects

Table I
Socio-demographic and clinical characteristics of the study group (*N*=304)

Age, years	
Mean (SD)	65.7 (9.3)
Range	50–82
Sex (%)	
Male	30
Female	70
Duration of OA, years	
Mean (SD)	9.5 (7.9)
Range	1–19
BMI	
Mean (SD)	33.8 (11.3)
Range	20.1–49.8
K/L rating score, no. (%)	
Grade 1	26 (8.5)
Grade 2	102 (33.5)
Grade 3	136 (44.8)
Grade 4	40 (13.2)
Number of comorbid conditions, no. (%)	
0	130 (42.8)
1	91 (29.9)
2	38 (12.5)
3	20 (6.6)
4 or more	25 (8.2)
Educational level, no. (%)	
Primary school	190 (62.5)
Secondary school	64 (21.0)
High school/university	50 (16.5)

BMI, body mass index (body weight divided by the square of the height).

Table II
Descriptive statistics and features of score distributions for health status measures in OA patients (*N*=304)

	Mean score	SD	Median	Interquartile (25th–75th)
WOMAC subscales				
Pain	195.7	110.4	196	101–277
Stiffness	77.4	51.2	70	30–119
Physical function	751.4	381.6	762.5	450–1050
WOMAC overall score	1025.8	519.8	1042.5	584–1412
SF-36 subscales				
Physical function	48.1	23.1	45	30–65
Role limitations (physical)	33.5	37.5	25	10–75
Body pain	37.5	17.5	41	22–50
Energy/vitality	50.4	20.6	50	35–67
Role limitation (emotional)	51.3	41.3	50	12–90
Mental health	58.2	26.1	60	40–78
Social function	60.1	24.8	62	40–87
General health perceptions	51.1	42.6	50	32–67
SF-36 PCS	40.8	22.8	39	24–58
SF-36 MCS	58.2	26.1	60	40–78
Lequesne algofunctional index	11.7	4.9	12	7–15

PCS, Physical Component Summary Scale Score; MCS, Mental Component Summary Scale Score.

enrolled, 174 (57.2%) reported one or more medical comorbidities, mostly cardiovascular (29.2%), respiratory (14.5%), and metabolic (11.5%) disorders. All subjects were affected by OA of the knee: 91% of them presented primary OA, with a radiological severity prevalently of second and third degree of Kellgren's scale[23] (34 and 44%, respectively).

DISTRIBUTION OF SCORES

Table II summarizes the mean, SD, median values, and interquartile for each of the aspects of health status covered by the WOMAC and SF-36, and for Lequesne algofunctional index. The distribution of the scores in the WOMAC (pain, stiffness, physical function, and overall scores) are presented in Fig. 1. The bar on the left of each graph represents the number of subjects with a score of 0 (floor effect); the bar on the right represents the number of subjects with a maximum possible score (ceiling effect). The WOMAC had negligible floor and ceiling effects in patients with OA of the knee.

RELIABILITY

Of the 304 patients enrolled in the study, 258 completed WOMAC twice within the stipulated 7 to 10 days with a mean of 8.7±2.1 days. The ICCs of the three dimensions pain, stiffness, and physical function were 0.86, 0.68, and 0.89, respectively. Cronbach's alpha coefficients were acceptable for all three dimensions of the WOMAC, according to standards recommended by Steiner and Norman[37]. Cronbach's alpha for the three dimensions pain, stiffness, and physical function were 0.91, 0.81, and 0.84, respectively.

CONSTRUCT VALIDITY

In testing for convergent validity between instruments (Table III), we found that correlation coefficients for the comparable dimension of the WOMAC and the SF-36 (pain and physical) were –0.611, and –0.706, respectively ($P<0.0001$). Figure 2 shows a scatter plot of patient's WOMAC function against the physical functioning score of the SF-36. For the overall scores, the Spearman correlation coefficient is –0.664 ($P<0.0001$). We also investigated the relationship of WOMAC scores with SF-36 MCS. As shown in Table III, the strongest correlations were between the WOMAC pain and physical scores and MCS (–0.584 and –0.567, respectively). The WOMAC subscales and overall score were highly (0.585 to 0.771; $P<0.0001$) correlated with Lequesne algofunctional index (Fig. 3). All three WOMAC dimensions correlated significantly with each other ($P<0.0001$). The strongest correlation was between physical function and pain (0.824; $P<0.0001$). No significant difference was observed in the comparison between the two subgroups of patients stratified by age (under 65 years vs 65 years or more), marital status (living together vs living alone), and BMI (BMI of 26–29 kg/m^2 vs BMI>30 kg/m^2). Women tended to report higher WOMAC score than men, but this was significant only for the WOMAC pain score (women: 211.1±112.1 vs men: 179.6±97.1; $P=0.04$). Also, WOMAC pain score, but not WOMAC stiffness and physical function scores was inversely correlated with years of formal education. Stratification into three categories confirmed that increasing education was associated with lower pain scores (WOMAC pain scores: primary school=206.2±111.1; secondary

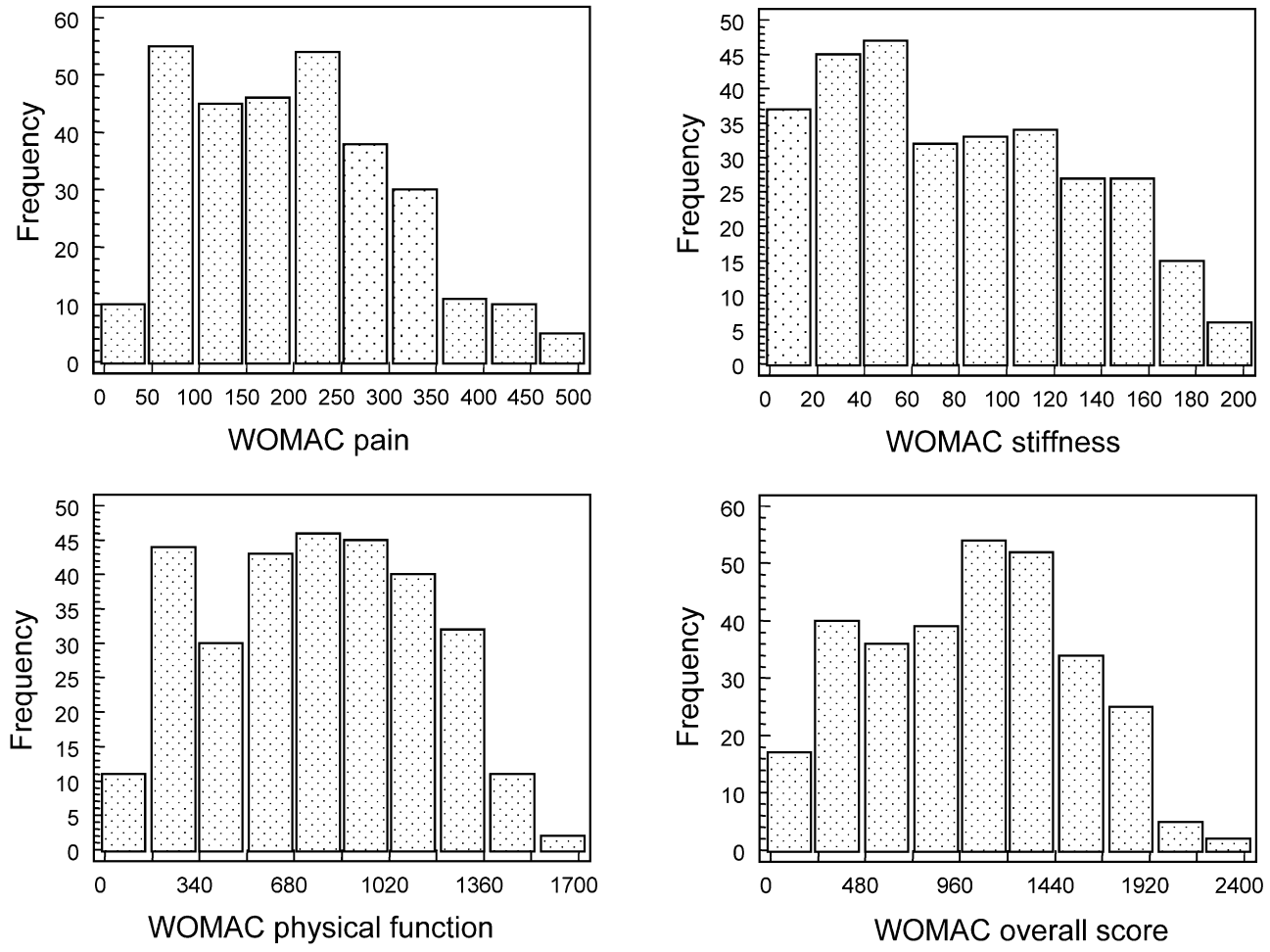

Fig. 1. Distribution of the scores in the WOMAC (pain, stiffness, physical function, and overall scores).

Table III
Convergent validity analysis: correlation matrix of WOMAC vs SF-36 dimensions and Lequesne algofunctional index (N=304)

	WOMAC pain	WOMAC stiffness	WOMAC physical function	WOMAC overall score
SF-36				
Body pain	–0.601	–0.462	–0.590	–0.608
Physical function	–0.611	–0.544	–0.706	–0.702
SF-36 MCS	–0.584	–0.403	–0.567	–0.572
SF-36 overall score	–0.608	–0.515	–0.650	–0.664
Lequesne algofunctional index	0.705	0.585	0.756	0.771

MCS, Mental Component Summary Scale Score.
All correlations were significant at P<0.0001 (Spearman rank coefficients).

school=189.4±112.2; high school/university=157.6±102.5; Kruskal–Wallis test: H=8.91; P=0.01) (Fig. 4). There were no significant differences in self-reported symptoms (pain and stiffness) by K/L scores, whereas WOMAC physical function dimension was weakly associated with radiological severity (K/L grade 1=609.2±349.2; grade 2=745.4±372.7; grade 3=728.1±357.1; grade 4=950.6±432.5; Kruskal–Wallis test: H=7.58; P=0.03).

DISCRIMINANT VALIDITY

Discriminant validity was assessed by comparing the WOMAC and SF-36 dimensions (pain, physical, and overall scores) in patients with and without other health conditions. For each dimension, the Spearman correlation coefficient was calculated to assess how well score correlate with the number of comorbid conditions. Positive correlations are expected in WOMAC, indicating that subjects with more comorbidities have higher scores than subjects with fewer comorbidities, while negative correlations are expected in SF-36, indicating that subjects with more comorbidities would have lower (worse) scores than subjects with fewer comorbidities. Correlations closer to 1 in absolute value indicate stronger correlation between the score and the number of comorbidities, and, therefore, more discriminatory ability. The SF-36 scores show a better

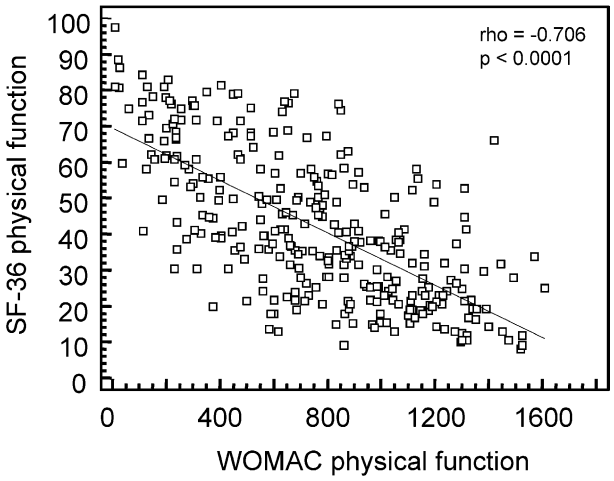

Fig. 2. Scatter plot of patient's WOMAC function against the physical functioning score of the SF-36. For descriptive linear regression line of 'best fit' purposes has been superimposed.

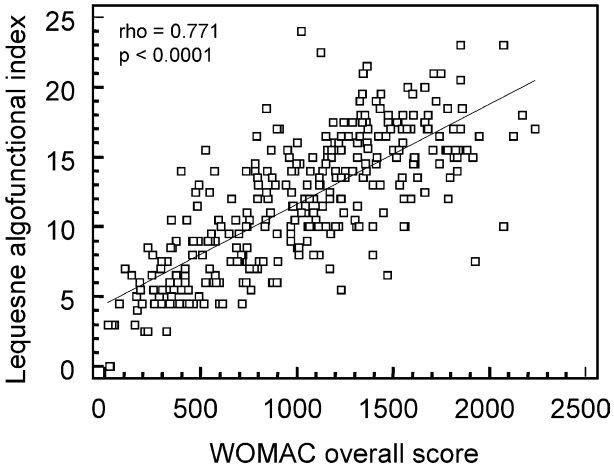

Fig. 3. Scatter plot of patient's overall WOMAC score against the score of the Lequesne algofunctional index. For descriptive linear regression line of 'best fit' purposes has been superimposed.

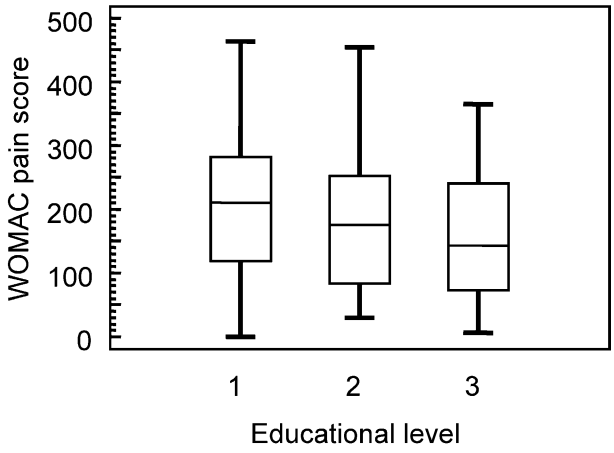

Fig. 4. Median WOMAC pain score by educational level (1, primary school; 2, secondary school; and 3, high school or university). The box plots provides information on the symmetry of a distribution, on the numerical measures of central tendency, and on the variability and spread of data in the tails of a distribution. The box contains the median values (represented by a horizontal line within the box), 25th and 75th percentiles, and wiskers representing the 10th and 90th percentiles. Kruskal–Wallis test was carried out across all three groups (P=0.01).

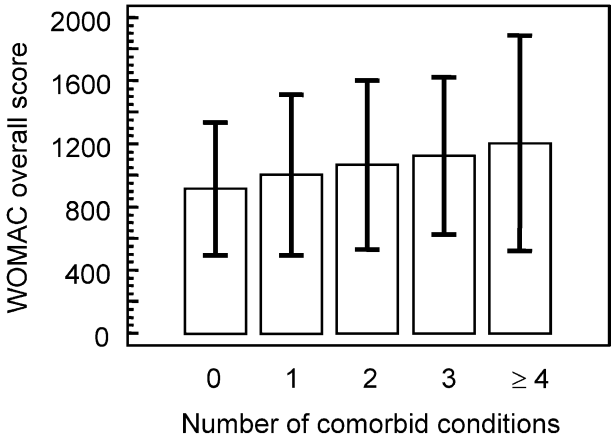

Fig. 5. Mean and SD of WOMAC overall scores by number of comorbidities.

gradient with comorbidities than the WOMAC. For pain scores, the Spearman correlation between number of comorbid conditions and the WOMAC and SF-36 scales, respectively, are 0.235 (P=0.0001) and –0.307 (P<0.0001). For the physical function score, the Spearman correlation between the number of comorbidities and the WOMAC and SF-36 are, respectively, 0.195 (P=0.0005) and –0.298 (P<0.0001). For the WOMAC and SF-36 overall scores, the Spearman correlation are, respectively, 0.211 (P=0.0003) and –0.312 (P<0.0001). Figures 5 and 6 shown the WOMAC and SF-36 overall scores by number of comorbidities.

Discussion

OA of the knee has been identified as one of the most prevalent chronic disorders affecting adults and a major cause of discomfort (pain and stiffness) and physical disability that results in extensive use of health-care resources[1–4,39]. In spite of the high prevalence of OA, presently, a few of validated health status measures exist for the evaluation of patients with OA, either in clinical practice or in clinical trials[40]. The WOMAC is a widely used and validated three-dimensional disease-specific, self-administered, health status measure assessing pain, stiffness, and function in patients with OA of the knee or hip[17,18]. Indicative of its widespread use in multicenter trials are its use by several groups of investigators and its incorporation in a set of guidelines for outcome measurement in trial of so-called slow acting drugs in OA (SADOA)[41], and in the core measures developed at the OMERACT III conference[19] and subsequently ratified by the OA Research Society International Task Force on

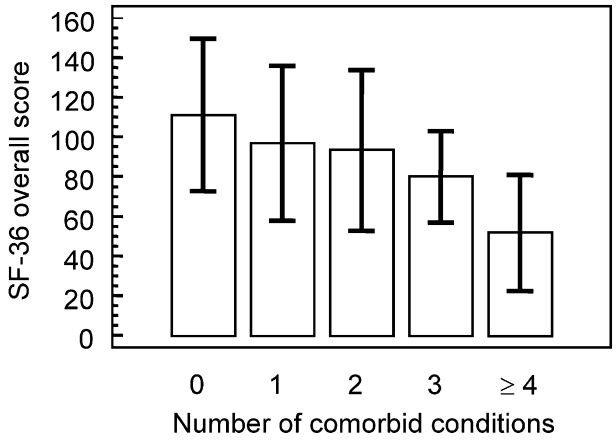

Fig. 6. Mean and SD of SF-36 overall scores by number of comorbidities.

clinical trials[40]. To enable comparison between assessments made in different countries, these measures need not only be translated, but also adapted for use in different cultures. The results of the present study show that the Italian version of WOMAC is a reliable and valid instrument for evaluating the severity of OA of the knee, with psychometric properties in agreement with the original widely used version.

Reliability was assessed in terms of internal consistency (Cronbach's alpha coefficient) and test–retest reliability (ICC analysis). Cronbach's alpha coefficients were acceptable for all three dimensions of the WOMAC, according to standards recommended by Steiner and Norman[37]; this indicates that each domain addressed a somewhat different aspect of functional disability. The ICCs of the WOMAC stiffness subscale was lower (0.68) than for pain (0.86) and physical function (0.89) subscales. This was expected, since similar findings have been reported for the original version[17,18], the German version[42], and the Swedish version[43]. For aggregate analysis, however, it is claimed that a reliability of >0.50 may be acceptable with a large sample size[44]. However, ICC analysis does result in lower values compared with the Pearson equivalent, as ICCs account for any additive or multiplicative element[45]. Furthermore, the WOMAC stiffness subscale is derived from only two scores, whilst the WOMAC pain and physical function scores are a mean of 5 and 17 VAS ratings, respectively. The inclusion of a greater number of subscales necessarily results in a more stable score, less susceptible to measurement error. This possibility partly explains the superior reliability demonstrated by the WOMAC pain and physical function dimensions[29]. Modest reliability may, however, just reflect fluctuating symptomatology characteristic of knee OA, as WOMAC asks the responders to consider their 'health state in the past 2 days'. In contrast, the SF-36 asks responders to consider 'the past 4 weeks', potentially allowing greater 1 week test–retest reliability compared with WOMAC[29].

Construct validity was examined in terms of the convergence between like dimensions of the self-administered WOMAC, and of the SF-36. A study of the correlation of like dimensions across the two health status instruments found the expected convergence. While WOMAC and SF-36 address symptoms and functional disability in separate scales, which may be aggregated into a composite index, the Lequesne OA algofunctional index directly aggregate symptoms and function, which are not graded separately. A very high correlation between WOMAC scores and Lequesne OA algofunctional index makes the Lequesne index redundant. In this study, as in others[21,24], a strong association was also noted between WOMAC subscale scores and patient's emotional state (SF-36 MCS). The fact that the WOMAC is sensitive to psychosocial factors is not to be attributed to the instrument itself. Self-report instruments is sensitive to these factors and, indeed, such factors contribute to the actual pain and physical impairment reported by patients[29,46–50]. If, however, a patient's emotional state markedly influence pain and physical health status perception, the resultant random measurement error would restrict the validity of the WOMAC or other self-report questionnaires to only relatively large studies[29]. In this study, we also investigated the relationship between WOMAC scores and the main sociodemographic characteristics (age, sex, disease duration, BMI, and educational level) and radiology K/L rating. No significant difference was observed in the comparison between the two subgroups of patients stratified by age (under 65 years vs 65 years or more). Previous researchers have failed to find an association between age and presence[48,51,52] or severity of pain and physical function in OA[46,47,53]. Female sex has been associated with increased reporting of knee pain in some community studies[50,52,53], but not in others[51]. Our results confirm earlier observations[50,52,53] that females tended to report greater severity of knee pain on WOMAC subscale. While BMI is clearly a strong risk factor for radiographic knee OA[6], its relationship with pain reporting is less certain[47,48,53]. We found that BMI was not associated with WOMAC dimension subscales. Years of formal education have been reported to be a risk factor for presence of knee pain in the community[53–55]. Previously[50], we found education to be related to mobility level, arm function, pain, and work as measured by Arthritis Impact Measurement Scales (AIMS2) subscores. We confirm a significant relationship between knee pain severity by WOMAC and level of formal education, suggesting that formal education should be included as a variable in clinical studies of knee OA. The mechanism by which education influences pain severity is unclear, but may be related to enhanced self-efficacy and sense of control allowing the patient to take advantage of a greater number of pain-reducing modalities. In this study, as in others[24,42,43,46–49,55–58], we found that radiographic severity as measured by K/L grade is not associated with pain severity. There are limitations to use radiographs for ascertainment of OA. Recent evidence suggests that radiographs may underestimate the true prevalence of OA[59]. This may result in misclassification of patients as to disease[6]. In addition, early disease may not be detectable by radiography, and some pathologic processes such as osteophytes may not represent progressive disease, thus potentially resulting in misclassification. Also, self-reported knee pain or physical function, which are a common complaint of elderly people, may not be due to a pathological process of the knee but to pain from hip or back disease, thus confounding the analyses[49]. Other prevalent causes of mobility restriction were the presence of problems with the cardiovascular and respiratory systems. Several studies, using data from the National Health Interview Survey Supplement on Aging[60–62] and Longitudinal Supplement on Aging[63], the Framingham Study[1], the Ontario Health Survey[64], and the Women's Health and Aging Study[65] have demonstrated the role of comorbidities

in the relationship between OA and disability. The results of the discriminant validity, assessed by comparing the WOMAC and SF-36 dimensions in patients with and without other health conditions pointed out that the SF-36 scores show a better gradient with comorbidities than that of the WOMAC on all three dimensions. This is not surprising since the WOMAC is designed as a measure of functional disability, rather than general health status, whereas, the SF-36 is a measure of general health status, which includes an assessment of functional disability and also assesses emotional functioning and roles, social functioning, and energy.

In conclusion, the results reported in this study confirm the reliability and validity of the Italian version of WOMAC in patients with OA of the knee. Collection information on health status using questionnaires such as WOMAC and SF-36 was acceptable to patients, though unfamiliar to them. Informally, patients reported a preference for the WOMAC because they found it easier to complete. Although we have not yet studied the sensitivity of WOMAC to change (i.e., responsiveness), this study has implications for the conduct of future clinical trials in OA. We are currently conducting further studies on the responsiveness of WOMAC against several other health status instruments.

Acknowledgements

The list of other GOQOLA investigators, given in alphabetical order, is as follows:
R. Bucci (U.O. di Reumatologia, OO.RR. Azienda Mista Ospedale Università di Foggia), M.A. Cimmino (DIMI, Università di Genova), A. Ciocci (Casa di cura 'Villa Domelia', U.O. di Reumatologia, Roma), F. Colombo (U.O. di Reumatologia, Ospedale Niguarda di Milano), L. Frizziero (U.O. di Reumatologia, Ospedale Maggiore, Bologna), V. Grattagliano (Cattedra e Divisione di Reumatologia, Azienda Ospedaliera Policlinico di Bari), M. Magarò (Istituto di Medicina Interna e Geriatria, Policlinico Gemelli, Università di Roma), M.T. Mascia (Cattedra e Servizio di Reumatologia, Policlinico di Modena), G. Minisola (U.O. di Reumatologia, Ospedale S. Camillo, Roma), V. Modena (U.O. di Reumatologia, Azienda Ospedaliera S. Giovanni Battista di Torino), M. Muratore (Servizio di Reumatologia, Ospedale G. Cascione, Lecce), A. Nervetti (U.O. di Reumatologia, Ospedale, Azienda di Parma), I. Olivieri (U.O. di Reumatologia, Ospedale S. Carlo, Potenza), G. Perpignano (I Cattedra di Reumatologia, Policlinico Universitario, Cagliari), S. Rossi (IRCCS Policlinico S. Matteo, Cattedra di Reumatologia, Università di Pavia), L. Sabadini (U.O. di Reumatologia, Azienda USL 8, Arezzo), P. Sarzi Puttini (U.O. di Reumatologia, Ospedale Sacco, Milano), F. Schiavon (Cattedra e Divisione di Reumatologia, Università di Padova), A. Spanò (Cattedra e Divisione di Reumatologia, Università Federico II di Napoli), A. Stancati (Clinica Reumatologica, Università di Ancona), S. Stisi (U.O. di Reumatologia, ASL 1, Benevento), F. Telese (Cattedra e Divisione di Reumatologia, Università di Ferrara), G. Tirri (Cattedra di Reumatologia, II Università degli Studi di Napoli).

The authors are very grateful to all the patients who willingly gave part of their time to complete the various assessments.

Appendix: The 24 items included in the Italian-WOMAC questionnaire

I. **DOLORE**

Che intensità prova:

1. Camminando su una superficie piana
2. Salendo o scendendo le scale
3. A letto, durante il sonno notturno (interferisce con il sonno)
4. Da seduto o in posizione supina
5. Stando in piedi, in posizione eretta

II. **RIGIDITA'**

Quanto è intensa la Sua rigidità:

6. Subito dopo il risveglio al mattino
7. Dopo essere stato seduto, sdraiato oppure dopo aver riposato, più tardi nel corso della giornata

III. **FUNZIONE FISICA**

Qual è il grado di difficoltà che avverte nel:

8. Scendere le scale
9. Salire le scale
10. Alzarsi da seduto
11. Stare in piedi
12. Piegarsi verso il pavimento (per raccogliere un oggetto)
13. Camminare su una superficie piana
14. Entrare ed uscire da una macchina o salire e scendere da un autobus
15. Andare a far spese
16. Mettersi i calzini o le calze
17. Alzarsi dal letto
18. Togliersi i calzini o le calze
19. Stare sdraiato a letto
20. Entrare ed uscire dalla vasca da bagno
21. Stare seduto
22. Sedersi o alzarsi dal water
23. Fare lavori domestici pesanti
24. Fare lavori domestici leggeri

References

1. Guccione AA, Felson DT, Anderson JJ, Anthony JM, Zhang Y, Wilson PW, *et al.* The effects of specific medical conditions on the functional limitations of elders in the Framingham Study. Am J Public Health 1994;84:351–8.
2. Yelin E. The economics of osteoarthritis. In: Brandt KD, Doherty M, Lohmander LS, Eds. Osteoarthritis. New York: Oxford University Press 1998;23–30.
3. Meenan RF, Callahan LF, Helmick CG. The National Arthritis Action Plan: a public health strategy for a looming epidemic Editorial. Arthritis Care Res 1999; 12:79–81.
4. Leardini G, Salaffi F, Montanelli R, Gertzeli S, Colangelo I, Canesi B. A multicentric study of annual costs of knee osteoarthritis in Italy. Arthritis Rheum 2001;44:S313.
5. Davis MA. Epidemiology of osteoarthritis. Clin Geriatr Med 1988;4:241–55.

6. Felson DT, Zhang Y. An update on the epidemiology of knee and hip osteoarthritis with a view to prevention. Arthritis Rheum 1998;41:1343–55.
7. Davis MA, Ettinger WH, Neuhaus JM, Mallon KP. Knee osteoarthritis and physical functioning: evidence from the NHANES I epidemiologic followup study. J Rheumatol 1991;18:591–8.
8. Liang MH, Fossel AH, Larson MG. Comparisons of five health status instruments for orthopedic evaluation. Med Care 1990;28:632–42.
9. Carr AJ. Beyond disability: measuring the social and personal consequences of osteoarthritis. Osteoarthritis Cartilage 1999;7:230–8.
10. Guyatt GH, Bombardier C, Tugwell PX. Measuring disease-specific quality of life in clinical trials. CMAJ 1986;134:889–95.
11. Patrick DL, Deyo RA. Generic and disease-specific measures in assessing health status and quality of life. Med Care 1989;27:S217–32.
12. Testa MA, Simonson DC. Assessment of quality of life outcomes. N Engl J Med 1996;334:835–40.
13. Ware JE Jr, Sherbourne CD. The MOS 36-item short form health survey (SF-36). 1. Conceptual framework and item selection. Med Care 1992;30:473–81.
14. Bergner N, Bobbitt RA, Carter WB, Gilson BS. The sickness impact profile: development and final revision of a health status measure. Med Care 1981; 19:787–805.
15. Hunt SM, McEven J, McKenna SP. Measuring health status: a new tool for clinicians and epidemiologists. J R Coll Gen Pract 1985;35:185–8.
16. Hurst NP, Jobanputra P, Hunter M, Lambert M, Lochhead A, Brown H. Validity of EuroQol—a generic health status instrument in patients with rheumatoid arthritis. Economic and Health Outcomes Research Group. Br J Rheumatol 1994;33:655–62.
17. Bellamy N, Watson Buchanan W, Goldsmith CH, Campbell J, Stitt LW. 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 the knee. J Rheumatol 1988;15:1833–40.
18. Bellamy N, Buchanan WW, Goldmith 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.
19. Bellamy N, Kirwan J, Boers M, Brooks P, Strand V, Tugwell P, *et al.* Recommendations for a core set outcome measure for future phase III clinical trials in knee, hip and hand OA. Consensus development at OMERACT III. J Rheumatol 1997;24:799–802.
20. Boers M, Brooks P, Strand VC, Tugwell P. The OMERACT filter for outcome measures in rheumatology Editorial. J Rheumatol 1998;25:198–9.
21. McConnell S, Kolopack P, Davis AM. The Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC): a review of its utility and measurement properties. Arthritis Care Res 2001;45:453–61.
22. Altman RD, Asch E, Bloch DA, Bole G, Borenstein D, Brandt K, *et al.* Development of 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.
23. Kellgren JH, Lawrence JS. Radiological assessment of osteoarthrosis. Ann Rheum Dis 1957;16:494–501.
24. Wolfe F. Determinants of WOMAC function, pain and stiffness scores: evidence for the role of low back pain, symptom counts, fatigue and depression in osteoarthritis, rheumatoid arthritis and fibromyalgia. Rheumatology 1999;38:355–61.
25. Apolone G, Mosconi P. The Italian SF-36 Health Survey: translation, validation and norming. J Clin Epidemiol 1998;51:1025–36.
26. Ware JE, Kosinski M, Keller SD. SF-36 Physical and Mental Health Summary Scales: A User's Manual. Boston: The Health Institute, New England Medical Centre 1994.
27. Ware J, Kosinski M, Bayliss M, Rogers WH, Raczec A. Comparison of methods for the scoring and statistical analysis of SF-36 health profile and summary measures: summary of results from the Medical Outcomes Study. Med Care 1995;4:AS264–79.
28. Brazier JE, Harper R, Munro J, Walters SJ, Snaith ML. Generic and condition-specific outcome measures for people with osteoarthritis of the knee. Rheumatology 1999;38:870–7.
29. Fransen M, Edmonds J. Reliability and validity of the EuroQol in patients with osteoarthritis of the knee. Rheumatology 1999;38:807–13.
30. Ruta DA, Hurst NP, Kind P, Hunter M, Stubbings A. Measuring health status in British patients with rheumatoid arthritis: reliability, validity and responsiveness of the short form 36-item health survey (SF-36). Br J Rheumatol 1998;37:425–36.
31. 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.
32. Lequesne MG, Samson M. Indices of severity in osteoarthritis for weight bearing joints. J Rheumatol 1991; 27:16–8.
33. Lequesne MG, Méry C, Samson M, Marty M. Comparison between the WOMAC and the Lequesne indices in patients with knee and hip osteoarthritis Letter. Osteoarthritis Cartilage 1998;6:141–2.
34. Steiner GL, Norman DR. Health Measurement Scales: A Practical Guide to their Development and Use, 2nd edn. Oxford: Oxford University Press 1996.
35. Nunnally JC. Psychometric Theory, 2nd edn. New York: McGraw-Hill 1978.
36. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951;16:297–334.
37. Steiner DL, Norman GR. Health Measurement Scales: A Practical Guide to their Development and Use. Oxford: Oxford University Press 1989.
38. Froberg D, Kane R. Methodology for measuring health state preferences. II: Scaling methods. J Clin Epidemiol 1989;42:459–71.
39. Gabriel SE, Crowson CS, O'Fallon WM. Costs of osteoarthritis: estimates from a geographically defined population. J Rheumatol 1995;43:23–5.
40. Osteoarthritis Research Society (OARS) Task Force Report: design and conduct of clinical trials in patients with osteoarthritis: recommendations from a task force on the Osteoarthritis Research Society. Osteoarthritis Cartilage 1996;4:217–43.
41. Lequesne M, Brandt K, Bellamy N, Moskowitz R, Menkes CJ, Pelletier JP, *et al.* Guidelines for testing slow acting drugs in osteoarthritis. J Rheumatol 1994;21:65–71.

42. Stuki G, Meier D, Stuki S, Michel BA, Tyndall AG, Dick W, *et al.* Evaluation of a German version of the WOMAC osteoarthritis index. Z Rheumatol 1996; 55:40–9.
43. Roos EM, Klassbo M, Lohmander LS. WOMAC osteoarthritis index. Reliability, validity, and responsiveness in patients with arthroscopically assessed osteoarthritis. Scand J Rheumatol 1999;28:210–5.
44. Brook R. Health Status Measurement, a Perspective on Change. London: Macmillan 1994 Chap 2.
45. McDowell I, Newell C. Measuring Health. A Guide to Rating Scales and Questionnaires, 2nd edn. New York: Oxford University Press 1996.
46. Summers MN, Haley WE, Reveille JO, Alarcon GS. Radiographic assessment and psychological variables as predictors of pain and functional impairment in osteoarthritis of the knee or hip. Arthritis Rheum 1988;31:204–9.
47. Salaffi F, Cavalieri F, Nolli M, Ferraccioli GF. Analysis of disability in knee osteoarthritis. Relationship with age and psychological variables but not with radiographic score. J Rheumatol 1991;18:1581–6.
48. Davis MA, Ettinger VH, Neuhaus JM, Barclay JD, Segal MR. Correlates of knee pain among United States adults with and without radiographic knee osteoarthritis. J Rheumatol 1992;19:1943–9.
49. Hopman-Rock M, Odding E, Hofman A, Kraaimaat FW, Bijlsma JWJ. Differences in health status of older adults with pain in the hip or knee only and with additional mobility restricting conditions. J Rheumatol 1997;24:2416–23.
50. Salaffi F, Piva S, Barreca C, Cacace E, Ciancio G, Leardini Gon behalf of Gonarthrosis and Quality of Life (GOQUOLA) Study Group. Validation of an Italian version of the arthritis impact measurement scales 2 (ITALIAN-AIMS2) for patients with osteoarthritis of the knee. Rheumatology 2000;39:720–6.
51. Hochberg MC, Lawrence RC, Everett DF, Cornoni-Huntley J. Epidemiological association of pain in osteoarthritis of the knee. Semin Arthritis Rheum 1989;18(Suppl 2):4–9.
52. Lethbridge-Cejku M, Scott WW Jr, Reichle R, Ettinger WH, Zonderman A, *et al.* Association of radiographic features of osteoarthritis of the knee with knee pain: data from the Baltimore Longitudinal Study of Aging. Arthritis Care Res 1995;8:182–8.
53. Creamer P, Lethbridge-Cejku M, Hochberg MC. Determinants of pain severity in knee osteoarthritis: effect of demographic and psychosocial variables using 3 pain measures. J Rheumatol 1999;26:1785–92.
54. Callahan LF, Smith WJ, Pincus T. Self-report questionnaires in five rheumatic diseases. Arthritis Care Res 1989;2:122–31.
55. Hannan MT, Anderson JJ, Pincus T, Felson DT. Educational attainment and osteoarthritis—differential associations with radiographic changes and symptom reporting. J Clin Epidemiol 1992;45:139–47.
56. Cicuttini FM, Baker J, Hart DJ, Spector TD. Association of pain with radiological changes in different compartments and views of the knee joint. Osteoarthritis Cartilage 1996;4:143–7.
57. Dekker J, Boot B, van der Woude LH, Bijlsma JW. Pain and disability in osteoarthritis: a review of biobehavioral mechanisms. J Behav Med 1992;15:189–214.
58. MPM Steultjens, Dekker J, Bijlsma WJ. Coping, pain, and disability in osteoarthritis: a longitudinal study. J Rheumatol 2001;28:1068–72.
59. Oliveria SA, Felson DT, Reed JI, Cirillo PA, Walker AM. Incidence of symptomatic hand, hip, and knee osteoarthritis among patients in a health maintenance organization. Arthritis Rheum 1995;38:1134–41.
60. Verbrugge LM, Cates DM, Ike RW. Risk factors for disability among U.S. adults with arthritis. J Clin Epidemiol 1991;44:167–82.
61. Verbrugge LM, Lepkowski JM, Konkol LL. Levels of disability among U.S. adults with arthritis. J Gerontol 1991;46:S71–83.
62. Guralnik JM, LaCroix AZ, Everett DF, Kovar MG. Aging in the Eighties: The Prevalence of Comorbidity and its Association with Disability Advance Data from Vital and Health Sciences, No. 170. Hyattsville: National Center for Health Statistics 1989 DHHS Publication No. PHS 89-1250.
63. Verbrugge LM. Disability transitions for older persons with arthritis. J Aging Health 1992;4:212–43.
64. Badley EM, Rasooly I, Webster GK. Relative importance of musculoskeletal disorders as a cause of chronic health problems, disability, and health care utilization: findings from the 1990 Ontario Health Survey. J Rheumatol 1994;21:505–14.
65. Hochberg MC, Kasper J, Williamson J, Skinner A, Fried LP. The contribution of osteoarthritis to disability: preliminary data from the women's health and aging study. J Rheumatol 1995;22(Suppl 43):16–8.