Clin Rheumatol (2015) 34:1955–1960
DOI 10.1007/s10067-015-2871-8

ORIGINAL ARTICLE

The correlation between radiographic knee OA and clinical symptoms—do we know everything?

Amir Herman · Ofir Chechik · Ganit Segal · Yona Kosashvili · Ran Lador · Moshe Salai · Amit Mor · Avi Elbaz · Amir Haim

Received: 27 October 2014 / Revised: 10 December 2014 / Accepted: 12 January 2015 / Published online: 22 January 2015

Abstract This study aims to evaluate the correlations between common clinical osteoarthritis (OA) diagnostic tools in order to determine the value of each. A secondary goal was to investigate the influence of gender differences on the findings. Five hundred and eighteen patients with knee OA were evaluated using the Western Ontario and McMaster Osteoarthritis Index (WOMAC) questionnaire, short form 36 (SF-36) Health Survey, and plain radiographs. Analysis of variance (ANOVA) was used to compare the different domains of the WOMAC and SF-36 questionnaires between genders and the radiographic scale. Higher knee OA x-ray grade were associated with worse clinical outcome: for women, higher scores for the WOMAC pain, function and final scores and lower scores in the SF-36 final score; in men, lower SF-36 overall and physical domains scores. Gender differences were found in all clinical scores that were tested, with women having worse clinical scores for similar radiographic grading (p values <0.001). Knee radiographs for OA have an important role in the clinical evaluation of the patient. Patients with higher levels of knee OA in x-ray have a higher probability of having a worse clinical score in the WOMAC and SF-36 scores. The gender differences suggest that for similar knee OA x-ray grade, women's clinical scores are lower.

Trial registration: NCT00767780

Keywords Function · Knee osteoarthritis · Pain · Quality of life · X-ray

A. Herman (✉)
Department of Orthopedic Surgery, Sheba Medical Center, Tel Hashomer, Israel
e-mail: amirherm@gmail.com

A. Herman
Talpiot Medical Leadership Program, Sheba Medical Center, Tel Hashomer, Israel

A. Herman
Sackler School of Medicine, Tel Aviv University, Tel Aviv, Israel

O. Chechik · R. Lador · M. Salai · A. Haim
Department of Orthopedic Surgery, Tel Aviv Sourasky Medical Center, Tel Aviv, Israel

G. Segal · A. Mor · A. Elbaz
AposTherapy Research Group, Herzliya, Israel

Y. Kosashvili
Department of Orthopedic Surgery, Rabin Medical Center, Petah Tikva, Israel

Background

Knee osteoarthritis (OA) affects about 10 % of the population over the age of 60, with increased prevalence among women and elderly patients [1–3]. Reliable grading of the severity of knee OA is important for monitoring the patients during their follow-up period and evaluation of various treatment modalities. Several clinical tools are currently in practice for objective and subjective assessment of symptoms and disease severity in knee OA [1]. At present, however, there is no consensus as for the "gold standard" for the evaluation of disease severity and progression [4]. Thus, it is difficult to assess the validity of the present clinical tools. Furthermore, several epidemiologic studies suggest a profound gender difference in the manifestations of symptoms and functional performance. Thus, while some clinical parameters could be valuable for pathology evaluation in females, others may be more suitable for males [5, 6]. Debi et al. concluded that males and females with knee OA adopt different gait strategies in response to knee OA symptoms. Furthermore, although there were no gender differences in radiographic grading of knee OA degeneration, some differences in self-assessment questionnaires

were noted [7]. The authors did not examine the correlation between the self-assessment questionnaires, and whether it changes in females/males, hence this data is missing.

Among the clinical tools to evaluate the severity of knee OA are the Western Ontario and McMaster Osteoarthritis Index (WOMAC) score, the Short Form 36 (SF-36) Healthy Survey, and plain x-ray. The WOMAC questionnaire is widely used for assessing and managing knee OA [8]. It has been extensively validated and is considered as a main index for evaluation of OA [8]. The SF-36 is a questionnaires designed to assess quality of life. It is used in a variety of illnesses and conditions and has previously been employed for studies evaluating knee OA [9]. Radiographic grading of knee OA is based on plain x-rays, mainly weight-bearing standing antero-posterior (AP) and lateral views. The Kellgren-Lawrence (KL) classification is often used to describe OA severity. According to the KL classification, formation of osteophytes, periarticular ossicles, narrowing of joint cartilage, pseudocystic areas with sclerotic walls and altered shape of bone ends are the hallmarks of this disease. Joints are graded as no changes (0), doubtful (1), minimal (2), moderate (3), or severe (4) [10]. The linear correlation between the radiographic findings and physical function in knee OA has been reported to be limited [3].

Other studies, using odds ratio, have shown that as compared to KL grade 0, knees with OA changes will perform worse in clinical questionnaires [11]. However, these approaches were limited in their analysis of data. They compared all KL grades to the baseline "no-change" grade which required dichotomizing the clinical results variable, e.g., pain vs no pain.

The goal of the current study was to examine the gender-specific relations between the clinical questionnaires' scores and radiographic changes in patients with primary knee OA. Contrary to other studies, we plan to compare all the radiographic knee OA grades to one another and establish order between all of them.

Materials and methods

The study cohort was comprised of consecutive patients with primary symptomatic knee OA [12] who were referred to the AposTherapy center between January 2009 and July 2010. Patients arriving to the therapy center undergo a standard documentation of the clinical and radiography data and complete self-assessment questionnaire. All data is stored in a secured database, which is listed under the Israeli database registry. The study was approved by the institutional ethic committee (no. NCT00767780).

Baseline WOMAC index and SF-36 health survey were completed by the patients and standard standing AP knee x-ray films were obtained prior to the initiation treatment at the center [12–14]. The x-ray films were done according to referral by the treating physician in several community-based medical centers. The digital x-ray films were copied to the AposTherapy center database and were analyzed by an experienced orthopedic surgeon to determine the osteoarthritic grade by Kellgren and Lawrence. According to Neogi et al., the inter-rater reliability for the Kellgren and Lawrence grade is 0.9 [15]. A search for eligible data was performed on the research database of the AposTherapy Center. Inclusion criteria were as follows: (1) Patients suffering from symptomatic knee OA for at least 6 months and who fulfilled the American College of Rheumatology (ACR) clinical criteria for OA of the knee [16], (2) Patients with a standing AP knee x-ray of both knees, and (3) Patients who completed the WOMAC questionnaire [12, 13] and SF-36 Health Survey [14]. Exclusion criteria included the following: (1) Neurological and rheumatic inflammatory diseases, (2) Corticosteroid injection within 3 months of the study, (3) Earlier knee surgery excluding arthroscopy, (4) Joint replacement of the hip or knee, (5) Instability of the knee due to traumatic ligament injury, and (6) OA in other lower extremity joints. All knee radiographs were graded in accordance with the KL grading system by a single experienced orthopedic doctor. Demographic and anthropometric parameters such as age, gender, and body mass index (BMI) were retrieved from the patient's medical record. Subjective pain, function, and quality of life perception were evaluated using the WOMAC questionnaire and SF-36 Health Survey. A validated Hebrew version of the WOMAC Index VAS scale questionnaire was used (version 3.1) [13]. This index contains 24 questions and is divided into three sections: pain (5 questions), stiffness (2 questions), and function (17 questions). The questions on the WOMAC score are each given a score on a 0- to 100-mm scale. The result is given as an arithmetic average of the relevant questions. Results range from 0 cm–100 mm, where 0 indicates no pain, stiffness, and functional limitation and 100 indicates the most severe pain, stiffness, and functional limitation. A valid Hebrew translation of the SF-36 original Likert scale was used to evaluate patient's quality of life [14]. The SF-36 health survey consists of 36 questions regarding physical and mental health status. Each question is given a grade between 0 and 100 and the score on each domain is presented as an arithmetic average of all relevant questions. The physical component summary (PCS) is a summary scale comprised of questions regarding physical health, role of limitation due to physical problems, pain, general health, and energy/fatigue. The mental component summary (MCS) is a summary scale comprised of questions in the fields of general health, energy/fatigue, social functioning, role of limitations due to emotional problems and emotional well-being. The results range between 0–100, where 0 indicates poor quality of life and 100 indicates excellent quality of life.

Statistical analysis

We used the worse of the two grades (left or right). The data was analyzed by an experienced biostatistician (A.H.). Categorical variables are presented as count (percent). Continuous variables are described as mean (standard deviation). Comparisons between categorical variables were done by the chi-square test.

Two-way analysis of variance model (ANOVA) was used to compare the clinical parameters between gender and OA radiographic grade. Each clinical parameter was used as a dependent covariate in the model while independent covariates were gender and OA radiographic grade. The interaction term between gender and osteoarthritis radiographic grade was not found to be statistically significant in all the models and so, it was omitted.

After comparing clinical parameters between men and women, additional analyses were performed separately for men and women. Hypothesis testing for stochastic orders in clinical parameter scores between radiographic were done. The Jonckheere-Terpstra [17, 18] was used to compare the distributions of clinical questionnaires between different radiographic grades of OA as described by KL.

This statistical method compared between several populations taking into account the different order of the populations in regards to one another. It tests the null hypothesis that all the distributions are equal to the hypothesis that the populations are stochastically ordered according to the knee OA grade. This is sometimes referred to as an order restricted hypothesis testing (see book by Silvapulle and Sen [19]).

A p value of <0.05 was considered as statistically significant. All reported p values are two-sided.

Results

Study population included 518 patients; of which 335 (64.7 %) were women and 183 (35.3 %) were men. The mean (standard deviation, SD) age was 63.5 (8.5) and 63.4 (10.4) for women and men, respectively. The average (SD) body mass index was 31.6 (7.1) and 30.6 (5.5) for women and med, respectively (p value=0.12).

The KL radiographic OA grade was scored between grades 1 and 4. The grade distribution in women was 57 (17 %), 118 (35.2 %), 96 (28.7 %), and 64 (19.1 %) patients in grades 1–4, respectively. The grade distribution in men was 26 (14.2 %), 60 (32.8 %), 66 (36.1 %), and 31 (16.9 %) patients in grades 1–4, respectively. There was no statistically significant difference in OA radiographic grade between men and women (p value=0.367).

The mean and standard deviation of each clinical questionnaire according to OA radiographical grade and gender is presented in Table 1. It can be seen that in all parameters, there is a statistically significant difference between genders. However, statistically significant between OA radiographic grade was found only in SF-36 overall score, WOMAC pain, function, and overall scores.

The median scores of the clinical questionnaires for women and men according to arthritis grade are presented in Tables 2 and 3, respectively. Table 2 shows that for women, a stochastic order exists between OA radiographic grades for WOMAC pain, function, and overall scores. The table shows higher scores (greater disability) as the radiographic OA grade increases. These results were statistically significant—p values of 0.024, 0.001, and 0.003, for WOMAC pain, function, and overall scores, respectively.

Table 1 Clinical parameters by gender and radiographic grade

		Grade 1	Grade 2	Grade 3	Grade 4	p value gender	p value grade
SF-36 pain	Male	56.1 (21.8)	56 (19.4)	52.8 (23)	46.5 (20.7)	0.001	0.205
	Female	44.7 (21.8)	43.5 (22.6)	40.9 (22)	40.8 (22)		
SF-36 overall	Male	60.1 (16.2)	60.8 (15.8)	58.2 (17.5)	51.6 (18.2)	0.001	0.018
	Female	53.1 (15.7)	52.4 (17.5)	49.2 (16.6)	47.8 (16.4)		
SF-36 PCS	Male	55.6 (17.6)	56.9 (16.3)	54.7 (19)	47.4 (18)	0.001	0.059
	Female	48.3 (15.4)	47.5 (18.1)	44.1 (16.5)	44.2 (16.9)		
SF-36 MCS	Male	66 (15.2)	68.1 (16.3)	65.4 (16.6)	62.2 (18.1)	0.001	0.675
	Female	61.7 (18.4)	58.9 (18.8)	58.9 (19.6)	58.9 (20.5)		
WOMAC pain	Male	38.8 (22.5)	30.7 (19.2)	38.2 (20.7)	40.8 (17.2)	0.001	0.027
	Female	42.6 (21.8)	45.7 (22.6)	48 (21.9)	51.7 (19.5)		
WOMAC stiffness	Male	40.1 (23.8)	31 (26.7)	43.1 (28.7)	36.5 (25.4)	0.001	0.464
	Female	45.1 (26.5)	50.1 (31.1)	49.8 (31.2)	51.8 (27)		
WOMAC function	Male	34.6 (19.8)	28.8 (18.9)	36.6 (21.6)	36.6 (19.2)	0.001	0.004
	Female	40.2 (20)	44.1 (23.5)	48.7 (21)	51.7 (20.2)		
WOMAC overall	Male	35.9 (18.5)	29.4 (18.2)	37.4 (20.6)	37.4 (17.6)	0.001	0.006
	Female	41.1 (19.2)	44.9 (22.1)	48.7 (20.7)	51.7 (19.5)		

p values were calculated using two-way ANOVA with main effects—gender and x-ray grade

Table 2 Clinical parameters by x-ray grade, women (N=335)

Radiographic grade	Grade 1 (N=57)	Grade 2 (N=118)	Grade 3 (N=96)	Grade 4 (N=64)	p value
SF-36 pain	45.0	45.0	45.0	42.5	0.313
SF-36 overall score	53.3	49.5	50.1	50.4	0.052
SF-36 PCS	47.8	44.6	42.4	41.7	0.097
SF-36 MCS	67.4	59.1	63.0	57.5	0.580
WOMAC pain	42.2	47.9	48.0	52.2	0.024
WOMAC stiffness	43.0	53.0	43.5	49.8	0.285
WOMAC function	39.6	44.4	48.2	55.3	0.001
WOMAC overall score	39.6	45.3	48.0	56.1	0.003

Median values are presented; p values were calculated using the Jonckheere-Terpstra test

For men, only the SF-36 overall score and SF-36 physical function scores showed stochastic orders between radiographic grade groups. These orders were found to be statistically significant (p value=0.032 and 0.047, respectively).

Figures 1 and 2 show data for WOMAC function and overall score according to radiographic grades for both genders. The boxplots serve to demonstrate the stochastic order relations between radiographic grades. The middle lines are the median (50th percentiles) while the boxes are the 1st and 3rd quartiles (25th and 75th percentiles, respectively). The lines and whiskers represent the minimum and maximum in each group. Focusing on Fig. 1 (women), it can be seen that there are patients in radiographic grade 1 that have a high WOMAC function score while some patients in grade 4 have a low score. However, most of the patients (the boxes themselves) in radiographic grades 1 and 4 have low and high WOMAC function scores, respectively. The boxplots for grades 2 and 3 are intermediate according to the grade. This difference between the groups was statistically significant (p value=0.003).

Table 3 Clinical parameters by radiographic grade, men (N=183)

Radiographic grade	Grade 1 (N=26)	Grade 2 (N=60)	Grade 3 (N=66)	Grade 4 (N=31)	p value
SF-36 pain	51.2	57.5	55.0	45.0	0.091
SF-36 general health	62.5	62.5	60.4	54.2	0.267
SF-36 overall score	59.6	60.4	55.4	52.9	0.032
SF-36 PCS	52.6	54.5	52.5	44.2	0.047
SF-36 MCS	69.0	71.8	66.4	63.6	0.234
W pain	39.1	25.0	36.3	40.4	0.090
W stiffness	40.5	25.8	43.2	37.0	0.375
W function	32.1	24.9	35.8	32.1	0.162
W overall score	34.4	26.0	38.2	31.9	0.121

Median values are presented; p values were calculated using the Jonckheere-Terpstra test

Discussion

The present study found that patients with higher structural severity of knee OA, as presented in radiographic KL grade, have worse clinical scores. This finding was statistically significant for women using the WOMAC pain, function, and overall scores. For men, statistically significant relations were found in the SF-36 physical and overall scores.

It should be emphasized that these relations are probabilistic relations, i.e., comparing two patients; the one with higher knee OA radiographic grade will have higher probability for low clinical function. This is not categorically true.

An important aspect of our finding is that for similar radiographic grade, women have worse clinical outcome scores. Furthermore, our results imply that the questionnaires themselves could be viewed as gender specific, i.e., the WOMAC score captures the differences in clinical outcome among women and the SF-36 among men. Such a result was not published elsewhere.

Creamer et al. studied the relations between clinical scores and radiographic knee OA grade [3]. They did not find correlation between the WOMAC score and the KL radiographic grade. However, it is important to note that they used linear correlation as their statistical method. Linear correlation means that a linear equation can be built between the KL grade and the clinical outcome [20]. Such a linear relation does not exist. We have shown that a probability (stochastic) order relations exists between the clinical and radiographic aspects of knee OA.

Duncan et al. and Neogi et al. have shown that the odds ratio for more severe clinical disease (e.g., higher WOMAC) is higher for high KL scores [11, 15]. The odds ratio is actually the ratio between two probabilities. This approach is similar to the approach we presented here. Drawbacks to the odds ratio approach is that a baseline radiographic grade needs to be established and every comparison is made to the baseline. In effect, Duncan et al. have shown that higher KL level has worse clinical outcome than the baseline normal KL level. In the present work, it was established that generally higher KL radiographic grades are associated in a worse clinical outcome.

This study has some limitations. Adding a sky view radiograph for the evaluation of the patellofemoral joint arthritis was found to increase the sensitivity of reontgenographic detection of knee OA, as patellofemoral involvement in addition to the medial or lateral compartments was found to be associated with increased pain [21]. Nevertheless, physicians still depend mostly on structural changes as reflected by anteroposterior radiographs when making treatment decisions. This was a cross sectional analysis of the correlations between the WOMAC and SF-36 questionnaires with radiographic assessments. Future studies should examine this correlation over time and see if structural changes detected by KL are

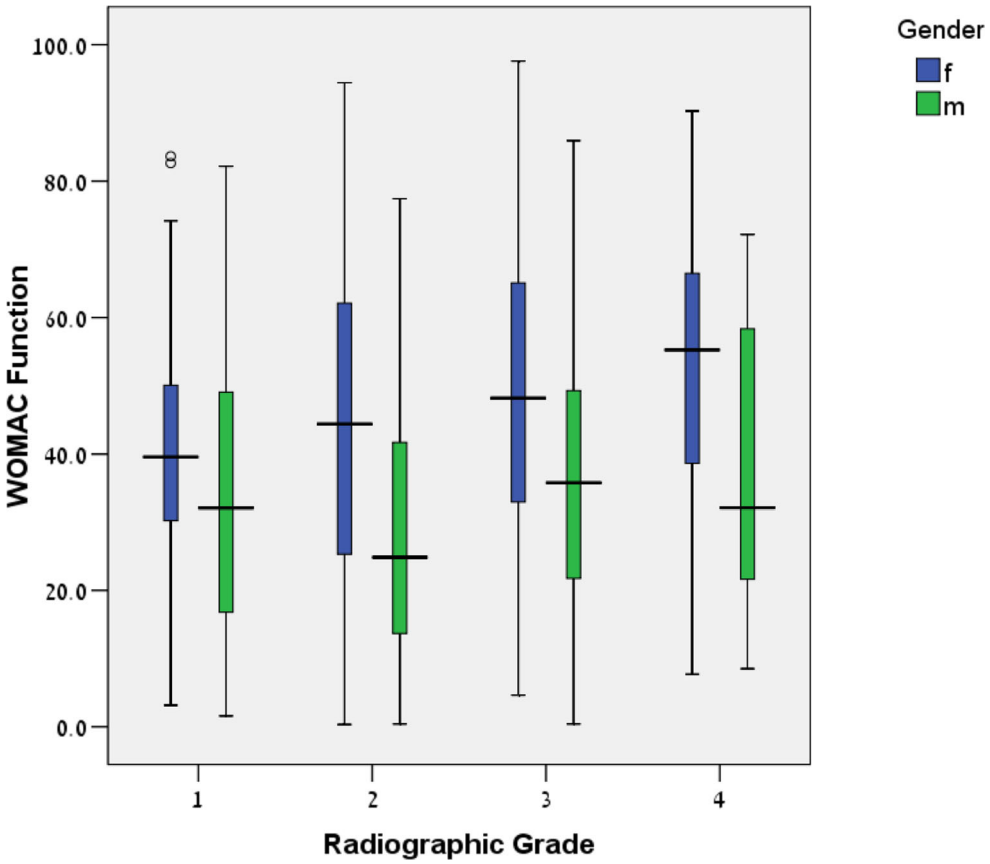

Fig. 1 Boxplots of WOMAC—function score. Boxplots of WOMAC function score: *middle line* represents the median, *boxes* are the first and third quartiles. The *whiskers* represent the minimum and maximum values. It can be seen that for female (f) as the radiographic grade increases the entire distribution (*box*) increase. This indicates a stochastically worse function as the radiographic grade increases, this was found to be statistically significant (*p* value=0.001). In men (m), this trend does not exists (*p* value=0.162)

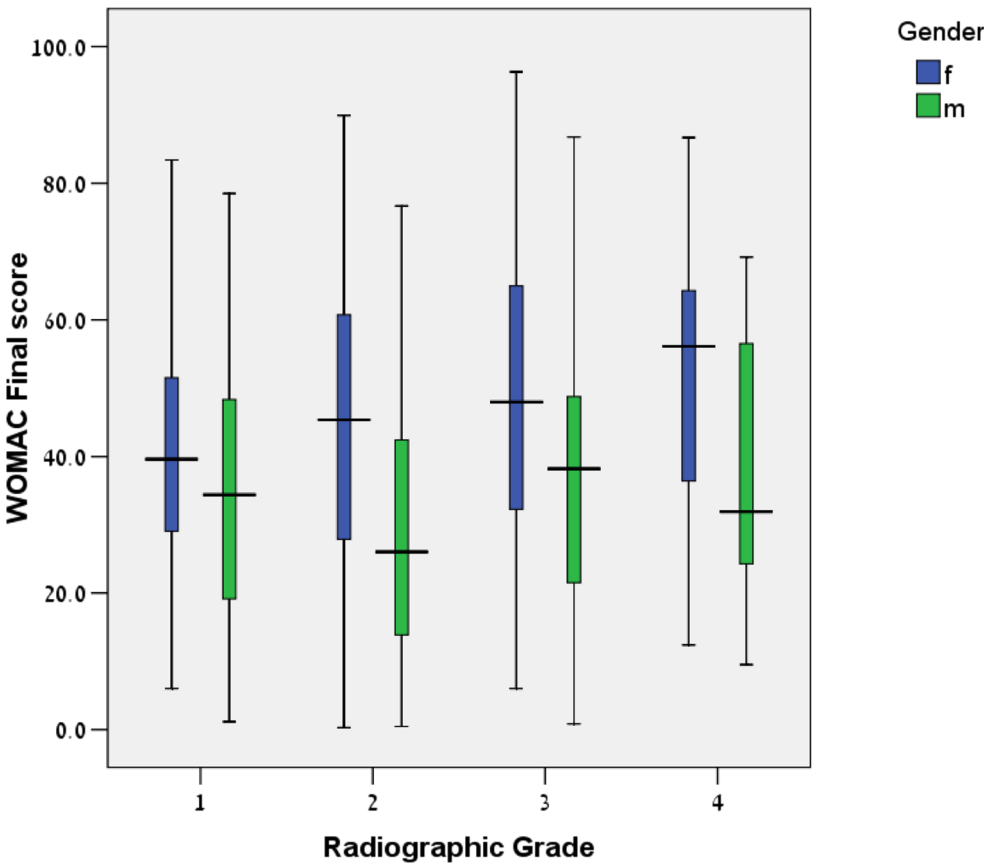

Fig. 2 Boxplots of WOMAC—overall score. Boxplots of WOMAC final score: *middle line* represents the median, *boxes* are the first and third quartiles. The *whiskers* represent the minimum and maximum values. It can be seen that for female (f), as the radiographic grade increases the entire distribution (*box*) increase. This indicates a stochastically worse function as the radiographic grade increases, this was found to be statistically significant (*p* value=0.003). In men (m), this trend does not exists (*p* value=0.121)

accompanied with higher levels of pain and poorer function and quality of life, and vice versa. Also, a longitudinal study may help to identify which questionnaire is more sensitive for symptomatic changes reported by the patients.

In conclusion, for patients with knee OA, the clinical state is probability ordered by the knee x-ray KL grade. Furthermore, significant differences were found between genders in their perception of pain and functional limitation.

Disclosures None.

Financial disclosures This study was not funded in any way.

Ethical approval The study was approved by the institutional ethic committee (No. NCT00767780). For this type of study, formal consent is not required.

References

1. Sun Y, Sturmer T, Gunther KP, Brenner H (1997) Reliability and validity of clinical outcome measurements of osteoarthritis of the hip and knee—a review of the literature. Clin Rheumatol 16:185–198
2. Rosemann T, Joos S, Koerner T, Szecsenyi J, Laux G (2006) Comparison of AIMS2-SF, WOMAC, x-ray and a global physician assessment in order to approach quality of life in patients suffering from osteoarthritis. BMC Musculoskelet Disord 7:6
3. Creamer P, Lethbridge-Cejku M, Hochberg MC (2000) Factors associated with functional impairment in symptomatic knee osteoarthritis. Rheumatology 39:490–496
4. Puhl W (1994) The importance of clinical subsets of osteoarthritis for epiderniological research. In: Puhl W, Brandt KD (eds) Epidemiology of Osteoarthritis. Georg Thieme Verlag Stuttgart, New York, pp 76–89
5. Badley EM (1995) The effect of osteoarthritis on disability and health care use in Canada. J Rheumatol Suppl 43:19–22
6. Dillon CF, Rasch EK, Gu Q, Hirsch R (2006) Prevalence of knee osteoarthritis in the United States: arthritis data from the Third National Health and Nutrition Examination Survey 1991–94. J Rheumatol 33:2271–2279
7. Debi R, Mor A, Segal O, Segal G, Debbi E, Agar G, Halperin N, Haim A, Elbaz A (2009) Differences in gait patterns, pain, function and quality of life between males and females with knee osteoarthritis: a clinical trial. BMC Musculoskelet Disord 10:127
8. Bellamy N (2002) WOMAC: a 20-year experiential review of a patient-centered self-reported health status questionnaire. J Rheumatol 29:2473–2476
9. Kosinski M, Keller SD, Ware JE Jr, Hatoum HT, Kong SX (1999) The SF-36 Health Survey as a generic outcome measure in clinical trials of patients with osteoarthritis and rheumatoid arthritis: relative validity of scales in relation to clinical measures of arthritis severity. Med Care 37(Suppl 5):MS23–MS39
10. Kellgren JH, Lawrence JS (1957) Radiological assessment of osteoarthrosis. Ann Rheum Dis 16:494–502
11. Duncan R, Peat G, Thomas E, Hay E, McCall I, Croft P (2007) Symptoms and radiographic osteoarthritis: not as discordant as they are made out to be? Ann Rheum Dis 66:86–91
12. Roos EM, Klassbo M, Lohmander LS (1999) WOMAC osteoarthritis index. Reliability, validity, and responsiveness in patients with arthroscopically assessed osteoarthritis. Western Ontario and MacMaster Universities. Scand J Rheumatol 28:210–215
13. Wigler I, Neumann L, Yaron M (1999) Validation study of a Hebrew version of WOMAC in patients with osteoarthritis of the knee. Clin Rheumatol 18:402–405
14. Lewin-Epstein N, Sagiv-Schifter T, Shabtai EL, Shmueli A (1998) Validation of the 36-item Short-Form Health Survey (Hebrew version) in the adult population of Israel. Med Care 36:1361–1370
15. Neogi T, Felson D, Niu J, Nevitt M, Lewis CE, Aliabadi P et al (2009) Association between radiographic features of knee osteoarthritis and pain: results from two cohort studies. BMJ 339:1–7
16. Altman R, Asch E, Bloch D et al (1986) 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 29:1039–1049
17. Jonckheere AR (1954) A distribution free k-sample test against ordered alternatives. Biometrika 41:133–145
18. Terpstra TJ (1952) The asymptotic normality and consistency of Kandall's test against trend when ties are present in one ranking. Indag Math 14:327–333
19. Silvapulle MJ, Sen PK (2005) Constrained statistical inference. John Wiley & Sons
20. Zou KH, Tuncali K, Silverman SG (2003) Correlation and simple linear regression. Radiology 227:617–628
21. Szebenyi B, Hollander AP, Dieppe P et al (2006) Associations between pain, function, and radiographic features in osteoarthritis of the knee. Arthritis Rheum 54:230–235