The Factor Subdimensions of the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) Help to Specify Hip and Knee Osteoarthritis.
A Prospective Evaluation and Validation Study

FELIX ANGST, THOMAS EWERT, SUSANNE LEHMANN, ANDRÉ AESCHLIMANN, and GEROLD STUCKI

ABSTRACT. ***Objective.*** To determine whether it is possible to specify different score patterns for hip and knee osteoarthritis (OA), and to identify the degree of responsiveness and the validity of the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) factors, which are alternative health dimensions obtained by factor analysis of the WOMAC items.

Methods. WOMAC scales and WOMAC factors were compared in a prospective setting examining patients with hip and knee OA before and after rehabilitative inpatient intervention (n = 317). In a partial sample (n = 103), the validity of the WOMAC factors was determined by a global rating of their activities.

Results. The WOMAC factor "ascending/descending" was significantly different for hip and knee OA in the health state before therapy (score in hip 5.09, in knee 6.59; p < 0.001); this was also true of the effect size after therapy (hip 0.39, knee 0.65; p = 0.012). The WOMAC scales did not differ for the 2 conditions. The WOMAC factor "ascending/descending" was the most responsive dimension in knee OA (effect size 0.65), but in hip OA the WOMAC pain scale was most responsive (effect size 0.55). Most of the WOMAC factors correlated moderately (r = 0.52–0.69) with the patient's self-rating on the validation questionnaire.

Conclusion. The WOMAC factors are valid measures. Analyzing the WOMAC by the WOMAC factors facilitates and improves the differential relevance and accuracy of the WOMAC for specific conditions such as hip and knee OA. (J Rheumatol 2005;32:1324–30)

Key Indexing Terms:

OSTEOARTHRITIS	OUTCOME	WOMAC FACTORS
EFFECT SIZE	RESPONSIVENESS	VALIDITY

Validated outcome assessment of quality of life and the quantitative evaluation methods of evidence based medicine are becoming increasingly important for a variety of reasons[1,2]. In particular, with regard to chronic, primarily incurable diseases such as osteoarthritis (OA), interest is not focused so much on the physical damage and its eradication but rather on the overall effects of the disorder and especially on functional health[2,3]. In addition to generic instruments for the comprehensive assessment of health, specific evaluatory instruments relevant to the health disorder are employed, e.g. the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC). The WOMAC is the most widely used self-rating quality of life instrument

for the lower extremity[1]. It is conventionally evaluated in 3 scales/dimensions, namely, (1) pain, (2) (joint) stiffness, and (3) function[4].

In order to improve the relevance and accuracy of the WOMAC with regard to specific symptoms and functional limitations, the construct similarities of the items were analyzed in 2 early studies using Rasch and factor analysis with the aim of differentiating various functional dimensions[5,6]. It was possible to identify 4 specific dimensions, in short, the WOMAC factors: (1) "lying/sitting," (2) "standing/walking," (3) "bending," and (4) "ascending/descending" (Table 1).

In a prospective, clinical intervention study, we investigated whether greater differentiation and sensitivity (responsiveness) could be achieved in the description of specific functional health related impairments of the lower extremity by working with the WOMAC factor dimensions rather than with the conventional WOMAC scale dimensions. Of particular interest was the question of whether hip and knee OA would produce different score patterns. This would be a very useful feature in demonstrating the success of a problem oriented, joint-specific intervention as is per-

Table 1. WOMAC scale and WOMAC factor dimensions.

	Items	Meaning
WOMAC scale		
Pain	P1–P5	Pain while walking on the flat, going up stairs, in bed, sitting or lying, standing upright
Stiffness	S1–S2	Stiffness after first awakening in the morning, later in the day
Function	F1–F17	Difficulties in descending, ascending, rising from sitting, standing, bending, walking, getting in/out of car, going shopping, putting on socks, rising from bed, taking off socks, getting in/out of bath, sitting, getting on/off the toilet, heavy domestic duties, light domestic duties
WOMAC factor		
Lying/sitting	P3	Pain while in bed at night
	P4	Pain while sitting or lying
	F12	Difficulty while lying in bed
	F14	Difficulty in sitting
Standing/walking	P1	Pain while walking on a surface
	P5	Pain while standing upright
	F4	Difficulty in standing
	F6	Difficulty in walking on flat
Bending	F5	Difficulty in bending to floor
	F9	Difficulty in putting on socks/stockings
	F11	Difficulty in taking off socks/stockings
Ascending/descending	P2	Pain while going up or down stairs
	F1	Difficulty in descending stairs
	F2	Difficulty in ascending stairs

formed in the field of rehabilitative medicine. The second aim of this study was to carry out a "subjective" (by the patient) and an "objective" (by the physician) validation of the factor dimensions.

MATERIALS AND METHODS

Study design and patients. Patients from the Zurzach osteoarthritis study[7,8], an ongoing clinical, prospective cohort study of patients with hip and knee OA, were evaluated in terms of the explanatory analysis of effects (see "evaluation study," Table 2). They were consecutively referred to the Rehaclinic Zurzach, Switzerland (formerly Zurzach Rheumatology and Rehabilitation Clinic) by a general practitioner or a rheumatologist and participated in a 3 week, inpatient, standardized rehabilitation program. The referring physician submitted a proof of failed outpatient treatment to the health insurance authority so that the costs for the clinic stay were paid for every patient. A screening log was created, recording all patients admitted for hip and knee OA and eligible for the study.

Inclusion criteria were: (1) patient fulfilled American College of Rheumatology (ACR) criteria for hip or knee OA, and (2) agreed to participate by written informed consent. Exclusion criteria were: (1) history of medication abuse, (2) severe illness (i.e., cancer, etc.), (3) German language skills insufficient to complete the questionnaires, (4) no wish to participate in the study. A detailed description of the sampling procedure and the intervention has been reported[8].

A subsample from this setting was collected consecutively during a certain period of the study and was additionally assessed with a view toward validation of the WOMAC factors ("validation study," Table 2). The assessments were carried out on admission (baseline) and discharge (followup) from the clinic.

Measures. The WOMAC is a self-rating instrument consisting of 24 items of which 5 (P1–P5) relate to pain, 2 (S1, S2) to joint stiffness, and 17 (F1–F17) to the function scale[4]. All items are rated on a numerical rating scale (Likert scale) of 0 (no symptoms/disability) to 10 (maximal symptoms/disability). The unweighted arithmetic mean of at least 4/5 pain, 1/2 joint stiffness, and 14/17 disability items make up the WOMAC scales, whereby 0 again represents the best and 10 the worst health condition. The derivation of the WOMAC scales and the WOMAC factors from the items is given in Table 1. The WOMAC factors are similarly determined by obtaining the unweighted arithmetic mean from 3 or 4 relevant items[5,6].

A validation questionnaire was especially compiled for our purposes and asked about the global function for each of the 4 factor dimensions. Both the patient (self-rating for internal validation purposes) and the treating physician (outside/external rating for external validation) completed the validation questionnaire. For example: "Please assess your present state of health in terms of complaints and limitations relating to the arthrosis when lying and sitting": no symptoms/limitations — maximal symptoms/limitations (not indicating the numbers). This rated symptoms/disability on an interval type scale from 0 = no symptoms/disabilities to 10 = maximal symptoms/disabilities. The same was done at the followup asking about the

Table 2. Patient characteristics.

	Evaluation Study			Validation Study		
	Hip	Knee	Total	Hip	Knee	Total
Total	126	191	317	27	76	103
Male (% per joint)	50 (40)	46 (24)	96 (30)	16 (59)	15 (20)	31 (30)
Female (% per joint)	76 (60)	145 (76)	221 (70)	11 (41)	61 (80)	72 (70)
Age, yrs, mean (SD)	64.4 (10.3)	66.1 (10.3)	65.4 (10.3)	66.5 (10.4)	68.1 (11.8)	67.7 (11.4)
Range	40.9–87.6	37.4–91.0	37.4–91.0	48.4–86.2	40.1–91.0	40.1–91.0

perceived response (by the patient or the observed change by the physician, respectively), as, for example, "Please assess the change of your health between entry in and discharge from the clinic in terms of complaints and limitations relating to the arthrosis when standing and walking" (analogously for the 3 other WOMAC factors): much worse, worse, slightly worse, equal, slightly better, better, much better (not indicating the numbers), i.e., 7 levels of a Likert scale. This is a measurement for validation purposes analogous to, for example, the global rating of pain (from 0 = no pain to 10 = maximal pain) for a pain scale (e.g., the WOMAC pain scale), which has been performed by various studies without specific pilot testing on reliability and validity of the global assessment.

Analysis. Changes in health status (effects) were determined by the effect size. The effect size is the difference in the score between the baseline (in our case, admission to the clinic) and the followup examination (discharge from the clinic) divided by the standard deviation of the baseline scores[9]. A positive effect size means improvement of health and vice versa. The effect size is simultaneously a measure of responsiveness (sensitivity to change) of an instrument's scale[7]. The difference between 2 corresponding baseline scores and the effect size for hip and knee OA patients was tested by the nonparametric Wilcoxon rank-sum test (Mann-Whitney U test) since some of the frequency distributions were not normally distributed (data not shown)[10]. Comparisons of the effect sizes of 2 health dimensions for the same patients were performed using the modified "jackknife test"[7,11].

The nonparametric Spearman rank correlation of the WOMAC factor score with the corresponding score of the validation questionnaire was defined as the measure of validity as previously defined in the WOMAC validation studies[4,10]. For the state at baseline, the global rating of the present state of health at entry to the clinic (from 0 = no symptoms/disabilities to 10 = maximal symptoms/disabilities) was correlated to the WOMAC factor score at baseline for each of the 4 WOMAC factors. For the effect at the followup, the global rating of the change of health between entry to and discharge from the clinic (1 = much worse to 7 = much better) was correlated to the change of the WOMAC factor score between baseline and followup for each of the 4 WOMAC factors. This was done for the patient (internal validation) as well as for the physician assessment (external validation). Correlation values of r = 0.50 to r < 0.71 were taken to be medium/moderate values indicating moderate clinical validity, because in this range 25%–50% (= r^2) of the variance of one variable can be explained by correlation with the other[12]. A value of r < 0.50 was considered a low, and r ≥ 0.71 a high correlation.

By sample size determination, assuming type I error α = 0.05 and power p = 0.80, the size of all strata (hip or knee OA) of the validation study subsample had to be at least n = 29 to obtain significant (Pearson) correlation coefficients r ≥ 0.50 explaining at least 25% (at least moderate correlation) of the variance of the WOMAC by the validation questionnaire[13].

All the analyses were performed using SPSS 11.0 for Windows (SPSS Inc., Chicago, IL, USA). Significance was accepted at the 5% level.

RESULTS

Patients. From September 18, 1996, to March 28, 2002, 317 consecutively referred patients were included in the setting of the evaluation study with their full set of data. Nine patients of the original 326 (2.8%) refused further participation at the followup. One hundred three (32.5%) patients were additionally assessed by means of the validation questionnaire (the validation study) between September 6, 2000, and March 28, 2002.

The baseline data are given in Table 2. The evaluation study included slightly more hip OA patients (40%) than the validation study (26%) and of the hip OA patients, more were women (60%) than in the validation study (41%). Age was comparably distributed between the 2 study sets (Table 2).

State of health and improvement in health after rehabilitation assessed by the WOMAC scale and the WOMAC factor dimensions (evaluation study). Table 3 shows the WOMAC baseline scores on admission to the clinic, the followup scores on discharge from the clinic, and the effect size at the end of the rehabilitation (followup) period, stratified for the hip and knee OA patients in the evaluation study (n = 317). The state of health on admission, reflected by the WOMAC baseline scores, shows that the knee OA patients reported slightly more symptoms than the hip OA patients, since their baseline scores were slightly higher on all 3 WOMAC scales than those of the hip OA patients. The WOMAC scales did not reveal any differences between hip and knee OA, although stiffness in the knee OA patients was, at 0.39 score points, slightly but not significantly worse (p = 0.112). The same applies to the WOMAC factors "lying/sitting" (p = 0.255) and "standing/walking" (p = 0.169). Pain and functional limitation in "bending" tended to be slightly more pronounced for the hip OA patients (difference, 0.62 score points; p = 0.061); however, in "ascending/descending," the knee OA patients reported significantly more symptoms (difference, 1.50 score points; p < 0.001). The same is true for the followup scores.

Improvements in state of health after the rehabilitative intervention, reflected by the effect size, were almost equal on the WOMAC scales for pain and function, whereby a small but not significant (p = 0.242) difference was observed for joint stiffness, with the knee OA patients experiencing slightly greater improvements (effect size 0.42) than the hip OA patients (effect size 0.28). For the 3 WOMAC factor dimensions "lying/sitting," "standing/walking," and "bending," the differences were also small and not significant. In "ascending/descending," however, the knee OA patients demonstrated an effect size (0.65) that was 67% higher, i.e., improvement in the state of health, than that of the hip OA patients (effect size 0.39, p = 0.012).

Responsiveness of the WOMAC factors compared with the WOMAC scales. The responsiveness of the WOMAC scales showed moderately high values of between 0.42 and 0.55 (with the exception of stiffness in hip OA: 0.28). The effect sizes of the WOMAC scales for pain and function were consistently higher than the effect sizes of the WOMAC factors. WOMAC pain, on the other hand, was significantly more responsive compared with all WOMAC factors, even compared with the most responsive factors "standing/walking" in hip OA (p < 0.001) and in knee OA (p = 0.006) (modified jackknife test). In contrast, only for the WOMAC factor "ascending/descending" was the responsiveness of the knee OA patients significantly higher than the WOMAC scales for pain (p = 0.006) and function (p < 0.001).

Correlation of the validation questionnaire to the WOMAC scores (validation study). The overall self-rating of limitations for the individual WOMAC factor dimensions with values between r = 0.62 and r = 0.68 correlated moderately

Table 3. WOMAC scores on admission (baseline), at discharge (followup), and effect sizes (ES) at discharge after inpatient rehabilitation (evaluation study: n = 126 hip OA, n = 191 knee OA patients).

| | Baseline Mean (SD) | | | Followup Mean (SD) | | | Effect Size (ES) | | |
	Hip	Knee	p	Hip	Knee	p	Hip	Knee	p
Pain	4.69 (2.15)	4.93 (2.13)	0.314	3.50 (2.25)	3.81 (2.27)	0.211	0.55	0.52	0.930
Stiffness	4.75 (2.51)	5.14 (2.56)	0.112	4.04 (2.46)	4.08 (2.41)	0.803	0.28	0.42	0.242
Function	4.66 (2.26)	4.92 (2.05)	0.284	3.71 (2.21)	4.02 (2.15)	0.186	0.42	0.44	0.526
Lying/sitting	4.14 (2.37)	3.82 (2.48)	0.225	3.18 (2.22)	3.07 (2.36)	0.463	0.40	0.30	0.197
Standing/walking	4.29 (2.41)	4.65 (2.32)	0.169	3.29 (2.33)	3.65 (2.44)	0.228	0.41	0.43	0.814
Bending	5.33 (2.85)	4.71 (2.56)	0.061	4.29 (2.73)	3.93 (2.53)	0.303	0.36	0.31	0.563
Ascending/descending	5.09 (2.71)	6.59 (2.27)	< 0.001	4.03 (2.54)	5.11 (2.47)	< 0.001	0.39	0.65	0.012

WOMAC scales: 0 = best, 10 = worst health. p values (hip vs knee) by Mann-Whitney U test.

to almost highly with the baseline score values before rehabilitative intervention in knee OA patients (Table 4). In hip OA patients, "ascending/descending" (r = 0.69) was almost highly correlated, "bending" moderately so (r = 0.53), and the 2 other WOMAC factors were consistently low (r ≤ 0.45). External rating of baseline health status by the physician produced poor correlations to the scores of equal or less than 0.47 for both joints.

Self-rating of improvement in state of health achieved after rehabilitation proved to be good for the hip OA patients according to the effect sizes (r = 0.61 to 0.62, except "bending": r = 0.43), whereas only "standing/walking" (r = 0.52) correlated moderately for the knee OA patients. The effect rating by the physician correlated only moderately for "standing/walking" (r = 0.50) and "ascending/descending" (r = 0.52) in the hip OA patients, but was low for the knee OA patients (r ≤ 0.39).

Comparison of male and female outcome. Post hoc analysis of the WOMAC data for male and female patients within the evaluation setting (n = 317) revealed that women's health status was consistently worse at baseline (higher WOMAC scores than the men; Appendix). This was especially the case in hip OA, where all differences attained "in trend" (0.050 ≤ p < 0.100) or significance (p < 0.050). In knee OA, only stiffness, function, and "ascending/ descending" revealed sex differences at baseline. At the followup examination, there were no sex differences. The effect sizes of the female patients were consistently higher (except in the WOMAC factor "bending") than those of the men, especially in stiffness, and less in "ascending/ descending."

Comparison of hip and knee OA within the sex strata showed that function measured by the WOMAC factor "ascending/descending" was significantly worse in knee OA in both sexes as well as at baseline and at followup, whereas improvement (as measured by effect sizes) in knee OA was higher only in the female patients. In both sexes, the WOMAC pain scale was the most responsive scale in hip OA (p = 0.098 in men, p = 0.002 in women), and the WOMAC factor "ascending/descending" was the most responsive dimension in knee OA (p = 0.238 in men, p = 0.013 in women; modified jackknife test). Sex-stratified analysis of the correlation data (validation setting) was not performed since 3 of the 4 strata were occupied by numbers of patients ≤ 16.

Table 4. Spearman rank correlation coefficients of the validation questionnaire scores with the WOMAC factor scores.

| | State at Baseline | | Effect at Followup | |
	Patient	Physician	Patient	Physician
Hip (n = 27)				
WOMAC factor				
Lying/sitting	0.35	0.12	0.62**	0.31
Standing/walking	0.45*	0.14	0.61**	0.50**
Bending	0.53**	0.47*	0.43*	0.41*
Ascending/descending	0.69**	0.18	0.61**	0.52**
Knee (n = 76)				
WOMAC factor				
Lying/sitting	0.65**	0.27*	0.20	0.25*
Standing/walking	0.68**	0.17	0.52**	0.39**
Bending	0.66**	0.09	0.17	0.14
Ascending/descending	0.62**	0.22	0.26*	0.11

** p < 0.01, * 0.01 ≤ p < 0.05 (2 tailed).

DISCUSSION

Specification properties and responsiveness of the WOMAC factors. Health status and its improvement (as measured by the effect size) after rehabilitation were equivalent for hip and knee OA patients if they were expressed by the WOMAC scales (except stiffness: higher effect sizes in female knee OA than in female hip OA). In contrast, the WOMAC factor dimension "ascending/descending" could clearly differentiate between the 2 conditions in both sexes: the baseline and followup scores as well as the effect sizes in the female patients were significantly different between hip and knee OA. In particular, the WOMAC pain and function scales revealed almost identical scores for hip and knee OA, also within both sex strata. Only for stiffness was a slight difference quantifiable, in that the female patients with knee OA reported a slightly higher level of improvement by the end of the clinic stay than the female hip OA patients (p = 0.054). The WOMAC factors "lying/sitting" and "bending" had slightly worse mean baseline scores in female hip OA.

Thus, it is possible to specify and differentiate hip and knee conditions with reference to state of health and alterations in it with reference to the WOMAC factors, at least and at best with the "ascending/descending" factor, but not with the WOMAC scales. This finding is in accord with daily clinical experience, which has shown that knee OA patients far more frequently report difficulties going up and down stairs than patients with hip conditions. However, it is the hip OA patients who have greater difficulty and more symptoms when performing functions involving hip flexion, such as bending (e.g., to tie shoes) and sitting, especially when sitting down in or getting up from an armchair. In rehabilitation medicine, the main symptomatic and functional problems experienced by the patient are recorded, and treatment is intensified and oriented specifically toward the main problem[2]. If the main problem can now be identified more sensitively and more specifically using the WOMAC factors, this will lead to a substantial improvement in the assessment of health status and changes in it, leading to optimal management of the rehabilitative interventions.

It is interesting that the WOMAC factors are now derived from items taken from the WOMAC scales. Thus, differentiation can be achieved with reference to some of the WOMAC pain and function items (P2 + F1 + F2 = WOMAC factor "ascending/descending"), but not by taking into account all the items that determine these 2 WOMAC scales. It can be assumed that the meaningfulness of the relevant items is lost in the background noise of the other 4 pain and 15 function items. Consideration of individual groups of items (i.e., the WOMAC factors) therefore produces a more differentiated picture of hip and knee conditions than the overall consideration of pain and function.

If the aim is to reflect changes in the health status as responsively as possible, only the WOMAC factor "ascend-ing/descending" appears to offer an advantage over the summarized WOMAC scales, and then only for patients with knee OA. In hip OA, the WOMAC pain was the most responsive dimension.

Validity of the WOMAC factors. The intensity of the symptoms and the loss of function for the individual dimensions of the WOMAC factors was, in most cases, generally reproducible for the patient, with a power to explain of 40%–50% (= r^2) of the variance of the scores. The validity of the construct of the WOMAC factors is thus confirmed by the patient's self-rating.

It is not confirmed when the external rating by the physician is regarded as the (gold) standard. This rating, with the exception of the effect sizes of individual factors for hip OA patients, correlated poorly with the results we obtained from the WOMAC questionnaire. However, we wanted to examine by the analysis of this external rating whether the state or the change of health can be reproduced by the clinical findings assessed by the physician. The answer is that it failed. The WOMAC per se is, however, a self-rating tool and it is questionable whether it could or should ever correlate with an external rating. There are numerous studies showing that self- and external ratings correlate poorly, especially with reference to OA[14,15]. Further, factor analysis was recently used to show that a medical rating — even if it is performed using standardized assessment instruments — remains a different evaluation construct from self-rating by a patient, even if, under ideal circumstances, a moderate correlation is observed[16].

The strength of this study is the impressive fact that the most important and most frequently employed instrument for the lower extremity, the WOMAC self-rating questionnaire, can be used in research as well as for daily clinical applications, and reveals very interesting and relevant findings and information. The differentiation between hip and knee conditions has, to our knowledge, not yet been achieved with any other self-rating instrument.

Another strength is the large size of the random sample recruited for the evaluation study from patients at a single clinic with a standardized intervention program, which gives the results a high statistical power. This also becomes apparent in that certain parameters, e.g., the age of the patients or some of the WOMAC scores (despite the closed scale of 0–10) and the effect sizes were of normal distribution. Despite some normally distributed parameters, nonparametric instead of parametric tests (e.g., t test) and Spearman rank instead of Pearson product-moment correlations were applied to maintain a conservative statistical strategy[10]. The relatively small size of the consecutively collected sample for the validation study is, however, simultaneously a weakness for the validation set, but one that plays a less important role as a possible source of bias when working with the Spearman method of nonparametric rank correlation. The size of the hip OA group was n = 27, making Pearson's r ≥

0.52 significant; i.e., the necessary n = 29 for r ≥ 0.50 was narrowly missed. This additionally inhibited post hoc analysis of the validation setting stratified by both sexes.

The relationship of the number of patients with hip/knee OA and the sex distribution of the hip OA patients differed in the evaluation and validation studies. On one hand, the risk of bias was eliminated by stratified analysis of hip and knee. On the other hand, a recent analysis indicated that sex was not a significant predictor for the magnitude of the effect size[17].

It is possible to question the quality and appropriateness of the questionnaire, given that the patient and the physician are equally required to make an overall assessment of abilities in the areas of lying/sitting, standing/walking, etc., in accord with the terms of the WOMAC factors, whereby these areas are very simply and generally expressed. It is undoubtedly not easy to complete the questionnaire, because the concept and understanding of the functional groups may differ greatly from one individual to the next.

We conclude the WOMAC factors are valid measures. Analyzing the WOMAC by the WOMAC factors facilitates and improves the differential relevance and accuracy of the WOMAC for specific conditions such as hip and knee, especially by the WOMAC factor "ascending/descending," which revealed significantly worse baseline and followup scores in knee OA than in hip OA for both sexes. In knee OA, the WOMAC factor "ascending/descending" was the most responsive dimension and more responsive than the WOMAC scales, especially in female patients. Whether the WOMAC factors will prove their worth in daily routine or whether additional differentiation of individual items will produce even better results requires investigation in future studies. Alternatively, studies examining item score patterns within hip and knee OA separately (by Rasch and/or factor analysis) could determine specific scores for hip and knee OA, with the possible disadvantage that different item patterns would not allow comparison of hip patients with knee patients.

ACKNOWLEDGMENT

REFERENCES

1. Angst F, Stucki G, Aeschlimann A. Quality of life assessment in osteoarthritis. Expert Rev Pharmacoeconomics Outcome Res 2003;3:623-36.
2. Stucki G, Kroeling P. Principles of rehabilitation. In: Hochberg MC, Silman AS, Smolen JS, Weinblatt ME, Weisman ME, editors. Rheumatology. 3rd ed. London: Mosby; 2003:11.1-11.14.
3. World Health Organization. International classification of functioning, disability and health. Geneva: World Health Organization; 2001:10-20.
4. Bellamy N. WOMAC Osteoarthritis Index. A user's guide. London,

APPENDIX. Post hoc analysis of WOMAC scores on admission (baseline), discharge (followup), and effect sizes (ES) at discharge after inpatient rehabilitation (evaluation study) stratified by sex.

| | Baseline Mean (SD) | | | Followup Mean (SD) | | | Effect Size (ES) | | |
	Male	Female	p	Male	Female	p	Male	Female	p
Hip (50 male, 76 female)									
Pain	4.23 (2.15)	4.99 (2.12)	0.075	3.29 (2.13)	3.64 (2.34)	0.380	0.44	0.64	0.216
Stiffness	4.22 (2.35)	5.10 (2.56)	0.065	3.81 (2.25)	4.20 (2.59)	0.504	0.17	0.35	0.436
Function	4.12 (2.14)	5.00 (2.29)	0.035	3.41 (2.13)	3.91 (2.26)	0.217	0.33	0.48	0.316
Lying/sitting	3.62 (2.17)	4.48 (2.44)	0.066	3.02 (2.08)	3.29 (2.32)	0.583	0.28	0.49	0.167
Standing/walking	3.77 (2.22)	4.63 (2.48)	0.057	3.05 (2.15)	3.45 (2.44)	0.338	0.32	0.48	0.447
Bending	5.02 (2.79)	5.53 (2.89)	0.031	4.07 (2.59)	4.43 (2.83)	0.546	0.34	0.38	0.776
Ascending/descending	4.59 (2.65)	5.42 (2.72)	0.102	3.75 (2.52)	4.22 (2.57)	0.325	0.32	0.44	0.832
Knee (46 male, 145 female)									
Pain	4.68 (2.20)	5.01 (2.11)	0.349	3.94 (2.43)	3.77 (2.22)	0.684	0.34	0.58	0.130
Stiffness	4.25 (2.68)	5.43 (2.46)	0.005	3.96 (2.63)	4.11 (2.34)	0.630	0.11	0.53	0.003
Function	4.38 (2.11)	5.10 (2.01)	0.026	3.66 (2.16)	4.13 (2.14)	0.237	0.34	0.48	0.332
Lying/sitting	3.58 (2.76)	3.90 (2.39)	0.303	3.06 (2.53)	3.08 (2.31)	0.714	0.19	0.34	0.232
Standing/walking	4.39 (2.15)	4.73 (2.38)	0.432	3.72 (2.48)	3.63 (2.44)	0.810	0.31	0.47	0.254
Bending	4.48 (2.46)	4.79 (2.60)	0.472	3.70 (2.63)	4.00 (2.50)	0.463	0.32	0.30	0.960
Ascending/descending	6.06 (2.22)	6.76 (2.27)	0.047	5.09 (2.47)	5.11 (2.47)	0.911	0.44	0.72	0.071
Hip vs knee: p values									
Pain	0.340	0.811		0.216	0.602		0.567	0.931	
Stiffness	0.988	0.256		0.843	0.972		0.499	0.054	
Function	0.540	0.817		0.590	0.436		0.692	0.790	
Lying/sitting	0.623	0.106		0.686	0.432		0.542	0.293	
Standing/walking	0.154	0.729		0.213	0.649		0.800	0.969	
Bending	0.361	0.065		0.532	0.335		0.806	0.535	
Ascending/descending	0.008	< 0.001		0.011	0.016		0.777	0.020	

WOMAC scales: 0 = best, 10 = worst health. p values (hip vs knee) by Mann-Whitney U test.

Canada: University of Western Ontario; 1995.

5. Ryser L, Wright BD, Aeschlimann A, Mariacher-Gehler S, Stucki G. A new look at the Western Ontario and McMaster Universities Osteoarthritis index using Rasch analysis. Arthritis Care Res 1999;12:331-5.

6. Ryser L, Aeschlimann A, Sangha O, Michel BA, Stucki G. Physical functional dimension scales are an alternative to condition-specific health status measures. Phys Med Rehab Kuror 2000;10:212-6.

7. Angst F, Aeschlimann A, Steiner W, Stucki G. Responsiveness of the WOMAC Osteoarthritis Index as compared with the SF-36 in patients with osteoarthritis of the legs undergoing a comprehensive rehabilitation intervention. Ann Rheum Dis 2001;60:834-40.

8. Weigl MB, Angst F, Stucki G, Lehmann S, Aeschlimann A. Long-term outcome in patients with osteoarthritis of the knee or hip after comprehensive rehabilitation: A prospective 2 year follow-up study. Ann Rheum Dis 2004;63:360-8.

9. Kazis ES, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. Med Care 1989;27 Suppl:S178-89.

10. Rosner B. The Wilcoxon rank-sum test. The correlation coefficient. In: Fundamentals of biostatistics. Boston: PWS-Kent; 2000:343-8,451-66.

11. Bessette L, Sangha O, Kuntz KM, et al. Comparative responsiveness of generic versus disease-specific and weighted versus unweighted health status measures in carpal tunnel syndrome. Med Care 1998;36:491-502.

12. Abrahamson JH, Abrahamson ZH. Making sense of data. A self-instruction manual on the interpretation of epidemiologic data. Oxford: Oxford University Press; 2001:200-1.

13. Bland M. Determination of sample size. In: An introduction to medical statistics. Oxford: Oxford Medical Publications; 1996:331-40.

14. Bruyere O, Honore A, Rovati LC, et al. Radiologic features poorly predict clinical outcomes in knee osteoarthritis. Scand J Rheumatol 2002;31:13-6.

15. Sun Y, Günther KP, Brenner H. Reliability of radiographic grading of osteoarthritis of the hip and knee. Scand J Rheumatol 1997;26:155-65.

16. Angst F, Pap G, Mannion AF, et al. Comprehensive assessment of clinical outcome and quality of life after total shoulder arthroplasty: Usefulness and validity of subjective outcome measures. Arthritis Rheum 2004;51:819-28.

17. Weigl MB, Angst F, Aeschlimann A, Lehmann S, Stucki G. Predictors for a good response to a comprehensive rehabilitation intervention in patients with OA of the knee or hip. Arch Phys Med Rehabil 2005; in review.