

Research

Open Access

Validation of the Rheumatoid and Arthritis Outcome Score (RAOS) for the lower extremity

Ann BI Bremander*¹, Ingemar F Petersson¹ and Ewa M Roos^{1,2}

Address: ¹Center for Research and Development, Spenshult Hospital for Rheumatic Diseases, SE-313 92 Oskarström, Sweden and ²Department of Orthopedics, Lund University Hospital, S-22185 Lund, Sweden

Email: Ann BI Bremander* - ann.bremander@spenshult.se; Ingemar F Petersson - ingemar.petersson@spenshult.se; Ewa M Roos - ewa.roos@ort.lu.se

* Corresponding author

Published: 17 October 2003

Received: 19 June 2003

Health and Quality of Life Outcomes 2003, 1:55

Accepted: 17 October 2003

This article is available from: <http://www.hqlo.com/content/1/1/55>

© 2003 Bremander et al; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

Abstract

Background: Patients with inflammatory joint diseases tend due to new treatments to be more physically active; something not taken into account by currently used outcome measures. The Rheumatoid and Arthritis Outcome Score (RAOS) is an adaptation of the Knee injury and Osteoarthritis Outcome Score (KOOS) and evaluates functional limitations of importance to physically active people with inflammatory joint diseases and problems from the lower extremities. The aim of the study was to test the RAOS for validity, reliability and responsiveness.

Methods: 119 in-patients with inflammatory joint disease (51% RA) admitted to multidisciplinary care, mean age 56 (± 13), 73% women, mean disease duration 18 (± 14) yr were consecutively enrolled. They all received the RAOS, the SF-36, the HAQ and four subscales of the AIMS2 twice during their stay for test of validity and responsiveness. Test-retest reliability of the RAOS questionnaire was calculated on 52 patients using the first or second administration and an additional mailed questionnaire.

Results: The RAOS met set criteria of reliability and validity. The random intraclass correlation coefficient (ICC_{2,1}) for the five subscales ranged from 0.76 to 0.92, indicating that individual comparisons were possible except for the subscale Sport and Recreation Function. Inter-item correlation measured by Cronbach's alpha ranged from 0.78 to 0.95. When measuring construct validity the highest correlations occurred between subscales intended to measure similar constructs. Change over time (24 (± 7) days) due to multidisciplinary care was significant for all subscales ($p < 0.001$). The effect sizes ranged from 0.30–0.44 and were considered small to medium. All the RAOS subscales were more responsive than the HAQ. Some of the SF-36 subscales and the AIMS2 subscales were more responsive than the RAOS subscales.

Conclusion: It is possible to adapt already existing outcome measures to assess other groups with musculoskeletal difficulties in the lower extremity. The RAOS is a reliable, valid and responsive outcome instrument for assessment of multidisciplinary care. To fully validate the RAOS further studies are needed in other populations.

Background

The rheumatic diseases include both inflammatory and non-inflammatory conditions. The chronic inflammatory diseases are all characterized by joint pain, joint swelling, morning stiffness, limitation of range of joint motion and in many cases a progressing physical impairment. The chronic inflammatory rheumatic diseases include a number of different diagnoses such as rheumatoid arthritis (RA), juvenile chronic polyarthritis and spondyloarthropathies [1].

Thanks to new treatments of inflammatory joint diseases the patients stay more alert and live a more active life compared to 10–20 years ago [2–4]. This change in physical status calls for assessment of items related to more difficult functions, such as sport and recreational activities. There are no self-administered questionnaires for lower limb function and chronic inflammatory joint diseases that take hip, knee and foot into account and at the same time relate to sport and recreational activities or to leg-related quality of life.

Functional disability and quality of life are key outcomes that influence the patients' compliance and satisfaction with the treatment and such measures should be based on self-assessment [5,6]. The Knee Injury and Osteoarthritis Outcome Score (KOOS) [7] is a self-administered extension of the WOMAC [8], and the validity, reliability and responsiveness has been found good in different populations with knee injuries and knee osteoarthritis [7,9–11]. The Foot and Ankle Outcome Score (FAOS) is an adaptation of the KOOS intended to evaluate symptoms and functional limitations related to the foot and ankle. The FAOS meet set criteria of validity and reliability [12]. The Hip Disability and Osteoarthritis Outcome Score (HOOS), another adaptation of the KOOS for people with hip osteoarthritis has also been shown to meet set criteria of validity and responsiveness [13]. The question was raised if the KOOS could be adapted and used to evaluate the outcome of patients with chronic inflammatory joint diseases and problems from the lower extremities.

The aim of the study was to test the reliability, validity and responsiveness of the Rheumatoid and Arthritis Outcome Score (RAOS), an adapted version of the KOOS, applied to people with chronic inflammatory joint disease and problems from the lower extremity.

Methods

An already existing questionnaire (Knee injury and Osteoarthritis Outcome Score, KOOS) was adapted for use in patients with inflammatory joint diseases by exchanging the word knee with leg in all the questions, no items were added or removed. The adapted questionnaire was called the Rheumatoid and Arthritis Outcome Score, RAOS.

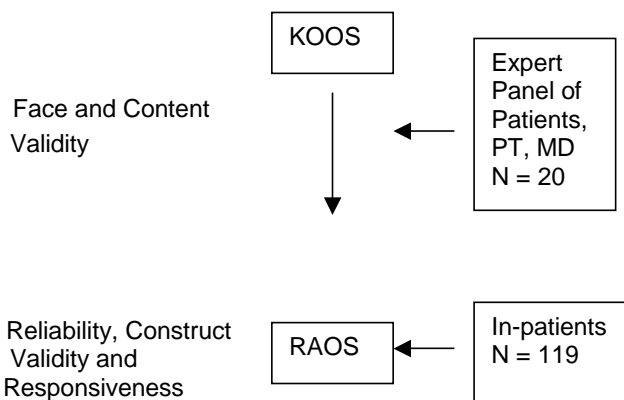


Figure 1
The adaptation and validation process of the Rheumatoid Arthritis Outcome Score (RAOS).

Firstly, the RAOS was reviewed by an expert panel to ensure face and content validity. Secondly the questionnaire was tested in a clinical study for assessing construct validity, reliability and responsiveness.

Expert panel

Thirteen patients with chronic inflammatory joint disease, 11 women and 2 men, mean age 56 (range 31 – 76), mean years of disease 14 (range 3.5 – 37), acted (after informed consent) as an expert panel to give the RAOS questionnaire face and content validity. Both in and outpatients were asked to participate, the emphasis put on a variety in age and years of disease. There was no set number of people who should be interviewed. The criterion 'sampling redundancy' was used; interviewing people until no new themes emerged [14] (Figure 1). In addition, two medical doctors and five physical therapists reviewed the questionnaire.

Development of the RAOS questionnaire

To assess content validity of the items the patients were asked to rate the relevance or importance of each item on a scale from one to three where: 1 = irrelevant, unimportant; 2 = somewhat relevant, somewhat important; 3 = very relevant, very important. It was considered that the mean score should be at least 2.0 (possible range 1.0 to 3.0) to justify inclusion into the RAOS. The same procedure was used when the KOOS was adapted for use in patients with problems related to the foot and ankle (the FAOS) [12]. The patients were asked to add items thought to be missing.

The expert panel added no items. However, due to difficulty acknowledging problems specifically related to the

Table 1: Patient characteristics

	Arthritis Total n = 119	Rheumatoid Arthritis n = 61	Inflammatory joint disease other than RA n = 58
Age in years; mean (SD)	56 (13)	61 (12)	50 (13)
Women; n, (%)	87 (73%)	47 (77%)	40 (69%)
Mean HAQ disability score; mean (SD)	1.3 (0.7)	1.4 (0.7)	1.1 (0.6)
Self-report of problem from the upper extremity %	100%	100%	100%
Self-report of problems from the lower extremity %	99%	100%	98%
Disease duration in years; mean (SD)	18 (14)	19 (14)	17 (14)
Days between questionnaires; mean (SD)	24 (7)	23 (5)	25 (8)

leg, the word was explained with hip, knee and foot where possible in the final RAOS questionnaire. All items had a relevance score over 2.0, the set criteria for inclusion in the RAOS, and no items were excluded because of poor content validity. The mean relevance score of all included items was 2.7 (range 2.4 – 3.0).

The RAOS questionnaire

The Rheumatoid and Arthritis Outcome Score (RAOS) is an adaptation of the Knee injury and Osteoarthritis Outcome Score (KOOS), intended to evaluate symptoms and functional limitations of importance to people with chronic inflammatory joint diseases and problems from lower extremities. The RAOS is a self-administered instrument and consists of 42 items assessing five separate patient-relevant dimensions: Pain (nine items); Other Symptoms like stiffness, swelling, and range of motion (seven items); Activities of Daily Living (ADL) (17 items); Sport and Recreational activities (Sport/Rec) (five items); and lower limb-related Quality of Life (QOL) (four items). The questions from the Western Ontario and Mac Master Universities (WOMAC) Osteoarthritis Index LK 3.0 [8] are included in their full and original form and WOMAC scores can thus be calculated from the RAOS questionnaire.

Five Likert-boxes were used (no, mild, moderate, severe, extreme) to answer each question. All items have a possible score from zero to four, and each of the five subscale scores is calculated as the sum of the items included. Raw scores are then transformed to a zero to 100, worst to best, scale. If a mark was placed outside a Likert-box the closest box was used. If two boxes were marked, the one indicating more severe problems was chosen. Missing data were treated as such; one or two missing values were substituted with the average value for the dimension. If more than two items were omitted, the response was considered invalid. The scores of the different subscales can be presented graphically as a RAOS profile. The RAOS question-

naire, user's guide and scoring manual can be downloaded from <http://www.koos.nu>.

Clinical study

A clinical study was designed to assess construct validity, reliability and responsiveness of the RAOS questionnaire. The study took place at Spenshult Hospital for Rheumatic Diseases, outside Halmstad in the southwest of Sweden.

Subjects

119 consecutively enrolled in-patients at Spenshult Hospital, mean age 56 (range 20 – 85), 73% women, mean disease duration 18 years (range 0.3 – 61), mean HAQ disability score 1.3 (range 0 – 2.88) were included in the study. Sixty-one of the patients were diagnosed with rheumatoid arthritis (RA) according to the ACR 1987 criteria [15]. The other 58 patients had an inflammatory joint disease other than RA; spondyloarthropathies (n = 24), polyarthritis (n = 4), psoriatic arthritis (n = 15), polymyalgia rheumatica (n = 2), Sjögren's syndrome (n = 5), Reiter's disease (n = 1), juvenile chronic arthritis (n = 6) and mixed connective tissue disease (n = 1) (Table 1). Patients undergoing post-operative rehabilitation were not asked to participate in the study.

All patients underwent exercise therapy and multidisciplinary team care during their stay at Spenshult. The physical training consisted of individual and group exercise led by a physical therapist.

Questionnaires

The SF-36 is a widely used generic instrument for assessment of health status. It is patient-administered and comprises eight subscales assessing physical and mental health to various degrees (Physical Function, Role-Physical, Bodily Pain, General Health, Vitality, Social Functioning, Role-Emotional and Mental Health) [16]. The Swedish acute version 1.0 was used [17].

The HAQ is a self-administered, disease-specific questionnaire. HAQ contains 20 items and assesses the degree of difficulty in performing activities of daily living during the last week. The activities are grouped into eight dimensions; Dressing and Grooming, Arising, Eating, Walking, Hygiene, Reach, Grip and Other Activities [18]. The HAQ is translated and validated for Swedish conditions [19].

The AIMS2 consists of 57 items. It can be divided into 12 scales: Mobility Level, Walking and Bending, Hand and Finger function, Arm Function, Self-Care Tasks, Household Tasks, Social Activities, Support from Family and Friends, Arthritis Pain, Work, Level of Tension and Mood. Together with questions about perceived current and future health and demographic data, it consists of a total of 78 items. The different subscales can be used solemnly [20,21].

Validity

The Short Form 36-item of the Medical Outcome Study (SF-36 acute version) [16], the Stanford Health Assessment Questionnaire (HAQ) [18], and four subscales of the Arthritis Impact Measurement Scale (AIMS2) [20] (Walking and Bending, Arm Function, Arthritis Pain, Level of Tension) were administered at baseline for determination of construct validity. High, medium or low correlations with the SF-36, the HAQ and the AIMS2 were hypothesized *a priori*. The highest correlations were expected when comparing scales intended to measure the same or similar constructs. We expected to observe higher correlations between the SF-36 Physical Function and the RAOS subscales ADL and Sport/Recreation than between SF-36 Role-Emotional and Mental Health compared to all the five RAOS subscales. The HAQ is a measure of ADL disability and were expected to have higher correlations to the RAOS subscale ADL than to the other RAOS subscales. For the AIMS2 the highest correlations were expected between Walking and Bending and RAOS subscale ADL and Sport/Recreation and also between AIMS2 Arthritis Pain and RAOS Pain. Lower correlations were expected between AIMS2 Arm Function and RAOS subscales Symptoms and Sport/Recreation since they do not measure similar constructs. Spearman's correlation coefficient (r_s) was used to assess construct validity [14,22]. When validating patient-relevant questionnaires, correlation coefficients between similar constructs often fall between 0.2 and 0.6 and rarely above 0.7 [23].

Floor and ceiling effects were assessed on the first administration of the RAOS for determination of content validity.

Reliability

To assess test-retest reliability, 67 of the enrolled patients had the RAOS questionnaire sent home, either prior to

admittance ($n = 17$) or after discharge ($n = 50$). Since no differences were seen in the results between these two groups, the results are reported for both groups together. A maximum of 15 days were allowed between the two assessments to minimize the influence of change in clinical status [14]. The test-retest reliability was calculated using the random-effects intraclass correlation coefficient ($ICC_{2,1}$) [14]. One suggestion for acceptable test-retest reliability for assessment of an individual is an intra class correlation coefficient of 0.85 [24]. When comparing groups, a lower intra class correlation coefficient is likely acceptable and a limit of 0.75 has been suggested [25].

According to Bland and Altman repeatability can be shown when plotting the difference against the mean of the two assessments for each subject. 95% of the differences are expected to be less than two standard deviations [26].

Internal consistency is an alternative approach to determine reliability, which is obtained from a single application of the technique and suggested because of the dynamic nature of many chronic diseases. A test with high inter-item correlation is homogenous and is likely to produce consistent responses [22]. Inter-item correlation was assessed on the baseline administration of the RAOS by calculation of Cronbach's alpha [14,22]. A Cronbach's alpha of ≥ 0.80 is generally regarded as acceptable [22].

Responsiveness

The patients completed the RAOS, the SF-36, the HAQ and four subscales of the AIMS2 at baseline and at the end of the multidisciplinary care intervention, shortly before leaving Spenshult. Change due to intervention, was assessed by Wilcoxon's signed rank test. Responsiveness was calculated by effect size. Effect size was defined as mean score change divided by the standard deviation of the baseline score [22]. Although there are no absolute standards for effect size it has been suggested that in comparative studies examples of small, medium, and large effect sizes might have values of 0.2, 0.5, 0.8, respectively [22].

Statistics

In the literature there is a lack of consensus on how to calculate reliability, validity and responsiveness of a questionnaire. The data obtained from questionnaires such as RAOS are ordinal and implies the use of non-parametric statistics. Statistical analyses of internal consistency have been made using parametric statistics as suggested by both Bellamy and Streiner [14,22]. The use of non-parametric statistics while checking for test retest reliability implies the use of the Kappa coefficient. If a quadratic weighting scheme is used, then the weighted kappa is exactly identical to the intraclass correlation coefficient (ICC) [14].

Table 2: Floor and ceiling effects of the questionnaires The percentage of patients reporting worst possible score (floor effect) / best possible score (ceiling effect) for the RAOS, the SF-36, the AIMS2 and the HAQ at baseline.

Questionnaire	Floor/Ceiling	Questionnaire	Floor/Ceiling
RAOS		SF-36	
Pain	0/3	Physical Function	4/1
Symptoms	0/2	Role Physical	64/10
ADL	0/2	Bodily Pain	7/1
Sport/Rec	37/1	General Health	2/1
QOL	3/1	Vitality	5/1
AIMS2 Walking / Bending	10/0	Social Functioning	1/21
Arm Function	1/18	Role Emotional	36/42
Arthritis Pain	5/1	Mental Health	1/4
Level of Tension	0/4	HAQ	0/3

Table 3: Construct validity Spearman's correlation coefficient (rs) determined when comparing RAOS five dimensions to the SF-36 eight different subscales, HAQ and four subscales of AIMS2. Significant correlations, $p < 0.05$ in bold figures, all correlations over 0.32 were significant at $p < 0.01$, $n = 115-119$.

	Reumatoid and Arthritis Outcome Score (RAOS)				
	Pain	Symptoms	ADL	Sport/Rec	QOL
SF-36					
Physical Function	0.56	0.47	0.65	0.63	0.53
Role Physical	0.31	0.18	0.38	0.18	0.26
Bodily Pain	0.57	0.34	0.56	0.21	0.46
General Health	0.35	0.27	0.35	0.20	0.33
Vitality	0.36	0.32	0.37	0.20	0.26
Social Functioning	0.34	0.21	0.37	0.12	0.27
Role Emotional	0.23	0.18	0.23	0.15	0.18
Mental Health	0.26	0.16	0.24	0.00	0.12
HAQ	-0.51	-0.38	-0.72	-0.64	-0.51
AIMS2					
Walking and Bending	-0.54	-0.47	-0.63	-0.48	-0.50
Arm Function	-0.38	-0.37	-0.46	-0.36	-0.36
Arthritis Pain	-0.50	-0.35	-0.45	-0.11	-0.42
Level of Tension	-0.23	-0.19	-0.17	-0.02	-0.15

Footnote: Negative correlations due to reversed scales for the HAQ and the AIMS2 (0 best) vs. RAOS and SF-36 (100 best).

Where parametric statistics have been used there were no differences between the results of parametric and non-parametric analyzes.

Analyses were carried out in both groups of patients (RA and other inflammatory joint diseases). Since the interpretation of the data was similar in both groups the results from all 119 patients will be reported together. Statistical significance was set to $p < 0.05$. The data was analyzed using SPSS 11.5.

The Ethics committee at the Medical Faculty at Lund University approved the study.

Results

Missing baseline data

For all study subjects, responses to sixty-four items were missing in the RAOS questionnaire (64 items of 42 items \times 119 patients = 1%). A total score could be calculated for all 119 subjects for the subscale Pain, for 118/119 subjects for the subscales Symptoms and Quality of Life, 117/119 subjects for Sport and Recreation and for the subscale ADL a total score could be calculated for 116/119 subjects.

Validity

Score distribution. The number of patients receiving floor or ceiling effects at baseline was low for the RAOS subscales, with one exception. For the subscale Sport/Recrea-

tion 43 subjects (37%) reported worst possible score (floor effect), indicating extreme problems with squatting, running, jumping, twisting/pivoting and kneeling at baseline. At follow-up the proportion reporting a floor effect decreased to 25%, indicating improvement occurring over time as measured by this subscale. All other RAOS subscales had little or no problem with floor or ceiling effects (Table 2).

When analyzed for construct validity the highest correlation occurred between subscales intended to measure similar construct, RAOS Activities of Daily Living vs. SF-36 Physical Function ($r_s = 0.65$) and RAOS Sport and Recreation vs. SF-36 Physical Function ($r_s = 0.63$). For the HAQ, the two highest correlations were HAQ vs. Activity of Daily Living ($r_s = -0.72$) and HAQ vs. Sport and Recreation ($r_s = -0.64$). Also for the AIMS2 the strongest correlation was Walking and Bending vs. Activity of Daily Living ($r_s = -0.63$), all correlations were significant at $p < 0.05$ (Table 3).

Reliability

67 questionnaires were sent home for test-retest reliability, 64 questionnaires were returned. Twelve subjects had to be excluded due to too long time elapsed (more than 15 days) between test and retest. For the remaining 52 subjects there was a mean of 9 days between test and retest (± 4 days). The random intraclass correlation coefficient ($ICC_{2,1}$) for the five subscales were Pain 0.87, Symptoms 0.85, ADL 0.92, Sport and Recreation 0.76 and for QOL 0.85. Bland and Altman plots of repeatability are given in Figure 2. For all subscales, 95% of the differences against the means were less than two standard deviations. Inter-item correlation, as measured by Cronbach's alpha was for the subscale Pain 0.92, Symptoms 0.78, ADL 0.95, Sport and Recreation 0.92 and for QOL 0.85.

Responsiveness

The mean number of days from baseline to follow-up was 24 days (range 12–58 days). A significant improvement was seen for all the RAOS subscales ($p < 0.001$) after the intervention multidisciplinary team care (Table 4).

The effect sizes for the five RAOS subscales were: Pain 0.40, Symptoms 0.41, ADL 0.44, Sport/Recreation 0.42 and QOL 0.30. Effect sizes for comparable subscales of the four different questionnaires are given in Figure 3.

Comparison of the RAOS to the SF-36, the HAQ and the AIMS2

Validity

When comparing the frequency of missing baseline data between the RAOS and the other three questionnaires used, the RAOS had the lowest percentage (1%) of missing values. For the SF-36, 134 items were missing (134 of

$36 \times 119 = 3\%$). Fifty-eight items were left out in the HAQ questionnaire (58 of $20 \times 119 = 2\%$) and in the AIMS2 70 items were lacking (70 items of $20 \times 119 = 3\%$).

37% of the patients reported worst possible score for the RAOS Sport and Recreation Function subscale. Other subscales with substantial floor and ceiling effects were SF-36 Role Physical (64%) and Role Emotional (36%) and AIMS2 Walking and Bending (10%). Substantial ceiling effects, indicating no possibility to measure improvement, were seen for AIMS arm function (18%), SF-36 Role Physical (10%), Social Functioning (21%) and Role Emotional (42%), Table 2.

Responsiveness

The effect sizes of the SF-36 ranged from 0.25 – 0.84, where the subscale Bodily Pain had a larger effect size than the corresponding subscale RAOS Pain. The SF-36 subscale vitality had the highest effect size (0.84) of all subscales in the study. The HAQ had a much smaller effect size than the RAOS subscale ADL supposed to measure similar constructs (0.14 vs. 0.44). The effect sizes of the AIMS2 ranged from 0.11 – 0.67, with Walking and Bending at the high end and Arthritis Pain at a medium effect size (0.43) comparable to the RAOS subscale Pain (Figure 3).

Discussion

The present and previous studies indicate that it is possible to adapt already existing outcome measures to assess similar groups of patients [8,12,13,27]. Developing an instrument is a time consuming process, effort and costs can be spared if already existing questionnaires can be adapted for use in similar groups of patients, assuming they meet set criteria. If a questionnaire is adapted to different areas and found to fulfill standard requirements it may be possible to make comparisons across diagnoses.

The RAOS has proven to be a reliable, valid and responsive outcome instrument for people with chronic inflammatory joint diseases and lower extremity dysfunction. The validation of an instrument is an ongoing process and testing validity arises not from a single powerful experiment, but from a series of converging experiments [14].

A questionnaire for the lower extremity

RA, and other inflammatory joint diseases, affects both the upper and lower extremity and to measure only lower extremity dysfunction could be questioned. There are however cases when the lower extremity is the key outcome area even if there are many areas of concern. For example interventions such as arthroplasty of the lower extremity or physical therapy treatment mainly directed towards the legs. Muscle dysfunction in the lower

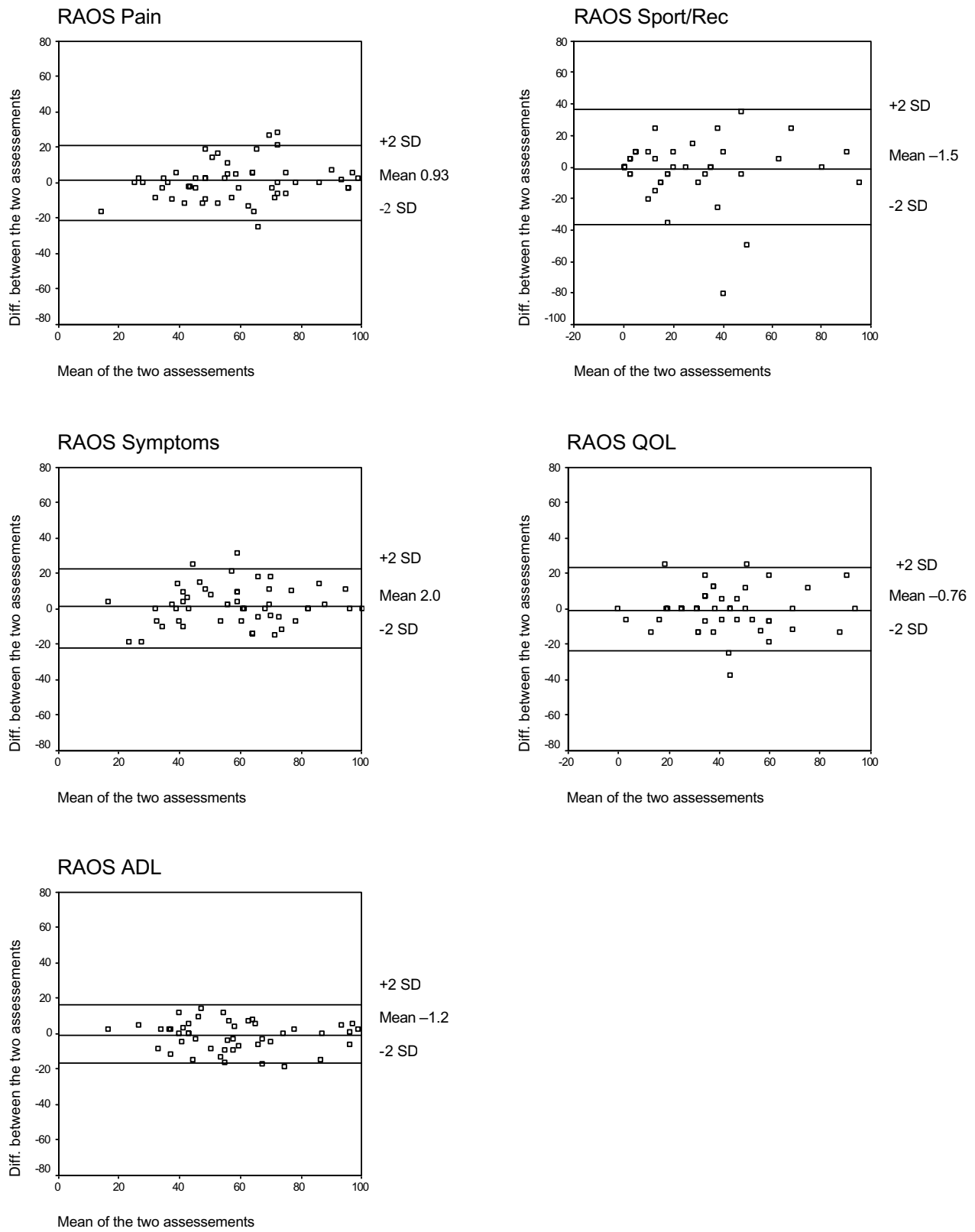


Figure 2
Bland and Altman plots for the five subscales.

Table 4: Mean (SD) of the RAOS at baseline and after the intervention multidiscipline care at Spenshult. 0–100 worst to best scale.

RAOS	Baseline	Follow up	P-value
Pain	51(21)	60(21)	<0.001
Symptoms	52(19)	60(20)	<0.001
ADL	50(21)	59(23)	<0.001
Sport/Rec	18(22)	27(29)	<0.001
QOL	35(21)	41(23)	<0.001

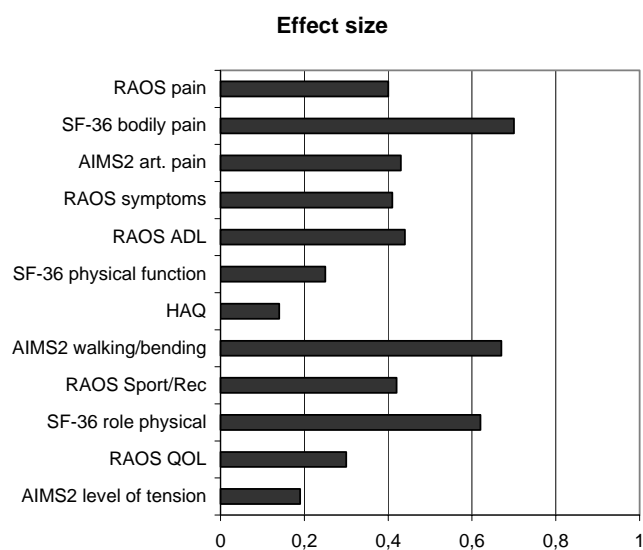


Figure 3
Effect size after intervention multidiscipline care for the five dimensions of the RAOS and corresponding dimensions of SF-36, HAQ and AIMS2.

extremity is common among people with inflammatory joint diseases [28,29]. A study by Ekdahl indicated that 80% of the patients with RA experienced muscle dysfunction in the lower extremities [30]. Commonly, outcome measures validated for RA focus on evaluating upper extremity dysfunctions. When an intervention is aiming at restoring lower extremity dysfunction, such an instrument is less valid and responsive and an outcome instrument validated for the lower extremity is a better choice. The RAOS is such an outcome measure. Also others have acknowledged the need for evaluation of lower extremity problems. Lately, some improvements to the HAQ have been made; introducing activities such as participation in sports and to do yard work [31], reflecting the need to evaluate more vigorous activities for people with inflammatory joint diseases.

Validity

Assessing validity is to measure the extent to which a technique measures that which is intended [22]. The expert panel rated the relevance of each item in the questionnaire, and found all original KOOS items being somewhat important or important. Another way of assessing content validity is to study the floor and ceiling effects of each item. A ceiling effect makes impossible measuring improvement while a floor effect makes impossible measuring deterioration. A low percentage of ceiling effects were seen for the RAOS indicating the RAOS having potential for measuring improvement over time. The number of patients having floor effects was small for all RAOS subscales except Sport/Recreation where 37% reported worst possible score at baseline. After intervention however the proportion of patients reporting floor effects was reduced to 25% indicating an improvement taking place and thus these functions being of importance to assess. This is in accordance with the opinion of the expert group who rated all the items in the subscale Sport and Recreation Function as important.

Generally it is well known that physical activity and physical function decline with age. To determine if older age or disease activity was associated with worse scores of the items included in the subscale Sport and Recreation Function we performed a logistic regression to analyze the risk of having a floor effect. A worse HAQ disability score ($p < 0.05$), but not older age, was associated with worse scores in items included in the subscale Sport and Recreation Function. This indicates that the subscale Sport and Recreation is as useful for patients of older age. This is in accordance with other validation studies of the KOOS; the subscale Sport and Recreation is as important to older patients with osteoarthritis as it is to younger individuals with osteoarthritis [11]. It has also previously been found that severe functional limitation affects this subscale more than it effects the other subscales of the KOOS [9].

Well-known and commonly used instruments for people with chronic inflammatory joint diseases were chosen to assess construct validity of the RAOS. In almost every study all over the world concerning arthritis and disability the HAQ is used and when studying health status the SF-36 is used. The AIMS2 is also commonly used; it consists of 12 different scales from which we choose four with the hypothesis high, medium or low agreement with the RAOS subscales. This is according to the suggestion of Liang and Jette; to fully establish construct validity the investigator must also demonstrate what variables are uncorrelated with the construct of interest [32]. In this study the correlations were as expected high when addressing subscales of similar construct and lower when compared to subscales assessing different constructs. However, to fully validate an outcome instrument it has to

perform as expected in different settings [14]. Further studies are needed to enlighten this question.

Reliability

Reliability is a measure of the consistency with which a technique yields the same results on repeated administration [14]. Test-retest was determined with a range of 1 – 15 days. The opinions regarding the appropriate interval vary from an hour to a year depending on the task, but a test-retest interval of two to 14 days is common for this type of questionnaire [14]. In our study one administration of the test-retest was given at home and the other one was given at the hospital. The difference in administration modes used may affect the reliability, but if so probably to the worse.

The test-retest reliability was high enough (ICC >0.85) to allow comparisons over time on an individual level for all subscales but Sport and Recreation (ICC 0.76) [24,25]. When studied in patients with knee injury, the Sport and Recreation Function subscale was more reliable (ICC 0.85) than in the present study, however compared to the other KOOS subscales it was less reliable [9]. Possibly a greater variability is to be expected when assessing more difficult physical function compared to activities of daily living and pain. In the revised and expanded version of the AIMS2 the test-retest reliability (ICC) for all 12 subscales ranged from 0.78 to 0.94, with a high correlation for the subscale Walking and Bending (ICC 0.92) [20].

To test for factors affecting the variability of the Sport and Recreation Function we checked for the impact of disease disability (HAQ score above median), older age, gender and disease duration. None of these factors were associated with increased variability of the subscale Sport and Recreation. However, it is well known that scales with more items have better reliability [14]. When comparing the Bland-Altman plots of the five RAOS subscales it is clearly seen that the fewer items of the subscale the worse test-retest reliability. To improve the reliability of the Sport and Recreation subscale, possibly items should be added. This strategy would however increase the length of the questionnaire. It should be determined if the reliability of the Sport and Recreation Function subscale could be improved to allow also comparisons on an individual level.

If the inventory of a questionnaire is relatively homogeneous and unambiguous, then the inter-item correlation will be high. The inter-item correlations of the five RAOS subscales were high enough to indicate homogeneity. According to Streiner a too high alpha (>0.90) may indicate item redundancy [14], which could be the case of the RAOS subscale ADL. The RAOS subscale ADL is equivalent to the WOMAC subscale Function. Item redundancy

for the WOMAC subscale Function has previously also been suggested by others [33].

Responsiveness

A small to medium effect size is to be expected when studying interventions such as multidimensional care and different forms of arthritis [21,34]. The value of HAQ as a group outcome measure is well established, the usefulness of monitoring individual HAQ scores in a clinical setting has been questioned [35]. A study by van den Ende et al. concluded that the HAQ is not an appropriate instrument to detect changes in physical impairments due to short-term exercise therapy [36], a finding confirmed by the current study. All the RAOS subscale scores improved significantly due to intervention and the RAOS effect sizes ranged from 0.30–0.44 indicating the RAOS being a valid measure for change over time. It is interesting to notice that the multidisciplinary care intervention improved not only difficulty with activities of daily living (as measured by the subscale ADL) but also improved more difficult physical functions (as measured by the subscale Sport and Recreation Function) to the same extent (as measured by similar effect sizes of the two subscales). By most other outcome measures this improvement would have remained undetected.

The multidisciplinary care given to the study subjects aimed at improving upper and lower extremity function, which could explain the generic SF-36 being more responsive than the RAOS with regard to the subscales Bodily Pain and Role Physical. Using multidisciplinary team care for validation of an instrument assessing only lower extremity function could be considered a limitation. It is thus of interest to note that the effect sizes of the RAOS were higher than for the HAQ, an instrument taking also other aspects into account and frequently used for assessment of multidisciplinary care in arthritis. The effect size of AIMS2 Arthritis Pain was of the same magnitude as the RAOS subscale Pain. The effect size of AIMS2 Walking and Bending was higher than of the corresponding subscales RAOS ADL and the SF-36 Physical Function. One possible explanation is the difference in response options in the questionnaires. The response alternatives in the AIMS2 are based on frequency of the difficulty and the SF-36 and the RAOS response alternatives concerns intensity of the difficulty. The subscale QOL was the least responsive of the RAOS instrument. As seen in studies on hip and knee replacement, this and similar subscales tends to need longer time to change than the 3–4 weeks in the present study [11,13].

Future application of the RAOS

Self-administered questionnaires can be generic or disease-specific. The advantage of using generic questionnaires such as SF-36 is that comparisons can be made

across diagnoses, and thus be a tool for health care planners. However, adapting a disease-specific questionnaire for musculoskeletal problems due to different origins could make comparisons across these diagnoses possible. The RAOS is such an adapted questionnaire also available for patients whose problems origin from the knee, hip and foot [7,12,13]. The RAOS includes the WOMAC, which is a widely used self-administered questionnaire for patients with osteoarthritis (OA) of either the hip or knee joint [8] validated also for patients with RA [27]. Adding dimensions such as Sport and Recreation Function and leg-related Quality of Life to the WOMAC can give a more descriptive picture of a subject, or a fuller picture of the impact of an intervention. We suggest using the RAOS to describe and follow patients with arthritis, especially when undergoing interventions aiming at restoring lower extremity function.

Conclusion

The present and previous studies indicate that it is possible to adapt already existing outcome measures to assess other groups of patients with musculoskeletal difficulties. The Rheumatoid and Arthritis Outcome Score (RAOS) is a reliable, valid and responsive outcome instrument for people with inflammatory joint diseases and lower extremity dysfunction undergoing a multidisciplinary care intervention. To fully establish the use of the RAOS questionnaire further studies are needed.

Authors' contributions

AB, ER and IP designed the study. AB collected the data, analyzed the data and drafted the manuscript. All three authors read and approved of the final manuscript.

Acknowledgements

Grants were obtained from the Research and Development Center of Spenshult, the Swedish Rheumatism Association, Lund University Medical Faculty and the Swedish Research Council.

References

- Benedek Thomas G: **History of the Rheumatic Diseases. Primer on the Rheumatic Diseases** 11th edition. Edited by: John H Klippel. Atlanta, Georgia, Arthritis Foundation; 1997.
- Geborek P, Crnkic M, Petersson IF and Saxne T: **Etanercept, infliximab, and leflunomide in established rheumatoid arthritis: clinical experience using a structured follow up programme in southern Sweden.** *Ann Rheum Dis* 2002, **61**:793-798.
- Jones G, Halbert J, Crotty M, Shanahan EM, Batterham M and Ahern M: **The effect of treatment on radiological progression in rheumatoid arthritis: a systematic review of randomized placebo-controlled trials.** *Rheumatology (Oxford)* 2003, **42**:6-13.
- Pincus T, Ferraccioli G, Sokka T, Larsen A, Rau R, Kushner I and Wolfe F: **Evidence from clinical trials and long-term observational studies that disease-modifying anti-rheumatic drugs slow radiographic progression in rheumatoid arthritis: updating a 1983 review.** *Rheumatology (Oxford)* 2002, **41**:1346-1356.
- Guillemin F: **Functional disability and quality-of-life assessment in clinical practice.** *Rheumatology (Oxford)* 2000, **39 Suppl 1**:17-23.
- Liang MH: **Longitudinal construct validity: establishment of clinical meaning in patient evaluative instruments.** *Med Care* 2000, **38**:1184-90.
- Roos EM, Roos HP, Lohmander LS, Ekdahl C and Beynon BD: **Knee Injury and Osteoarthritis Outcome Score (KOOS)--development of a self-administered outcome measure.** *J Orthop Sports Phys Ther* 1998, **28**:88-96.
- Bellamy N, Buchanan WW, Goldsmith CH, Campbell J and 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 knee.** *J Rheumatol* 1988, **15**:1833-1840.
- Roos EM, Roos HP, Ekdahl C and Lohmander LS: **Knee injury and Osteoarthritis Outcome Score (KOOS)--validation of a Swedish version.** *Scand J Med Sci Sports* 1998, **8**:439-448.
- Roos EM, Roos HP and Lohmander LS: **WOMAC Osteoarthritis Index--additional dimensions for use in subjects with post-traumatic osteoarthritis of the knee. Western Ontario and MacMaster Universities.** *Osteoarthritis Cartilage* 1999, **7**:216-221.
- Roos EM and Toksvig-Larsen S: **Knee injury and Osteoarthritis Outcome Score (KOOS) - validation and comparison to the WOMAC in total knee replacement.** *Health Qual Life Outcomes* 2003, **1**:17.
- Roos EM, Brandsson S and Karlsson J: **Validation of the foot and ankle outcome score for ankle ligament reconstruction.** *Foot Ankle Int* 2001, **22**:788-794.
- Nilsdotter AK, Lohmander LS, Klassbo M and Roos EM: **Hip disability and osteoarthritis outcome score (HOOS)--validity and responsiveness in total hip replacement.** *BMC Musculoskeletal Disord* 2003, **4**:10.
- Streiner DL Norman GR: **Health Measurement Scales. A practical guide to their development and use.** New York, Oxford University Press; 1995.
- Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, Healey LA, Kaplan SR, Liang MH, Luthra HS, Medsger TA, Mitchell DM, Neustadt DH, Pinals RS, Schaller JG, Sharp JT, Wilder RL and Hunder GG: **The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis.** *Arthritis Rheum* 1988, **31**:315-324.
- Ware J. E., Jr. and Sherbourne CD: **The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection.** *Med Care* 1992, **30**:473-483.
- Sullivan M Karlsson J: **Hälsöenkät: Svensk Manual och Tolkningsguide (Swedish Manual and Interpretation Guide).** Gothenburg, Health Care Unit, Sahlgrenska Hospital, Sweden; 1994.
- Fries JF, Spitz P, Kraines RG and Holman HR: **Measurement of patient outcome in arthritis.** *Arthritis Rheum* 1980, **23**:137-145.
- Ekdahl C, Eberhardt K, Andersson SI and Svensson B: **Assessing disability in patients with rheumatoid arthritis. Use of a Swedish version of the Stanford Health Assessment Questionnaire.** *Scand J Rheumatol* 1988, **17**:263-271.
- Meenan RF, Mason JH, Anderson JJ, Guccione AA and Kazis LE: **AIMS2. The content and properties of a revised and expanded Arthritis Impact Measurement Scales Health Status Questionnaire.** *Arthritis Rheum* 1992, **35**:1-10.
- Archenholtz B and Bjelle A: **Reliability, validity, and sensitivity of a Swedish version of the revised and expanded Arthritis Impact Measurement Scales (AIMS2).** *J Rheumatol* 1997, **24**:1370-1377.
- Bellamy N: **Musculoskeletal Clinical Metrology.** Dordrecht, Kluwer Academic Publishers Group; 1993.
- McDowell I Newell C: **measuring health: A guide to rating scales and questionnaires.** New York, Oxford University Press; 1987:27-231.
- Weiner E Stewart B: **Assessing individuals: Psychologic and educational tests and measurements.** Boston, Little Brown; 1984.
- Rosner B: **Fundamentals of Biostatistics.** Belmont, CA, Duxbury Press; 1995.
- Bland JM and Altman DG: **Statistical methods for assessing agreement between two methods of clinical measurement.** *Lancet* 1986, **1**:307-310.
- Wolfe F and Kong SX: **Rasch analysis of the Western Ontario MacMaster questionnaire (WOMAC) in 2205 patients with osteoarthritis, rheumatoid arthritis, and fibromyalgia.** *Ann Rheum Dis* 1999, **58**:563-568.

28. Bearne LM, Scott DL and Hurley MV: **Exercise can reverse quadriceps sensorimotor dysfunction that is associated with rheumatoid arthritis without exacerbating disease activity.** *Rheumatology (Oxford)* 2002, **41**:157-166.
29. Hakkinen A, Haanonan P, Nyman K and Hakkinen K: **Aerobic and neuromuscular performance capacity of physically active females with early or long-term rheumatoid arthritis compared to matched healthy women.** *Scand J Rheumatol* 2002, **31**:345-350.
30. Ekdahl C, Andersson SI and Svensson B: **Muscle function of the lower extremities in rheumatoid arthritis and osteoarthritis. A descriptive study of patients in a primary health care district.** *J Clin Epidemiol* 1989, **42**:947-954.
31. Pincus T, Swearingen C and Wolfe F: **Toward a multidimensional Health Assessment Questionnaire (MDHAQ): assessment of advanced activities of daily living and psychological status in the patient-friendly health assessment questionnaire format.** *Arthritis Rheum* 1999, **42**:2220-2230.
32. Liang MH and Jette AM: **Measuring functional ability in chronic arthritis: a critical review.** *Arthritis Rheum* 1981, **24**:80-86.
33. Ryser L, Wright BD, Aeschlimann A, Mariacher-Gehler S and Stucki G: **A new look at the Western Ontario and McMaster Universities Osteoarthritis Index using Rasch analysis.** *Arthritis Care Res* 1999, **12**:331-335.
34. Angst F, Aeschlimann A, Steiner W and 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**:834-840.
35. Greenwood MC, Doyle DV and Ensor M: **Does the Stanford Health Assessment Questionnaire have potential as a monitoring tool for subjects with rheumatoid arthritis?** *Ann Rheum Dis* 2001, **60**:344-348.
36. van den Ende CH, Breedveld FC, Dijkmans BA and Hazes JM: **The limited value of the Health Assessment Questionnaire as an outcome measure in short term exercise trials.** *J Rheumatol* 1997, **24**:1972-1977.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

