The Responsiveness of Generic Quality of Life Instruments in Rheumatic Diseases. A Systematic Review of Randomized Controlled Trials

ZULMA ORTIZ, BEVERLEY SHEA, MARCELO GARCIA DIEGUEZ, MAARTEN BOERS, PETER TUGWELL, ANELUS BOONEN, and GEORGE WELLS

ABSTRACT. To review the available evidence that has used generic instruments alone or in comparison with disease specific instruments. A systematic review was carried out using the methods recommended by the Cochrane Collaboration. We used MEDLINE and EMBASE searches and we performed a hand search of the abstracts listed under "quality of life" at American College of Rheumatology (ACR) meetings. Selection was limited to randomized controlled trials (RCT) using generic instruments in populations older than 18 years with any of the following diseases: rheumatoid arthritis, fibromyalgia, osteoporosis, osteoarthritis, systemic lupus erythematosus, and ankylosing spondylitis. Language was restricted to English papers. Studies using only disease-specific instruments were excluded. From 488 articles retrieved, 13 reports of 10 randomized controlled trials were selected. There were 101 abstracts on quality of life in ACR abstract books; 78 abstracts contained data on generic instruments, and of these, 9 described their use in RCT. Despite a substantial increase in the number of papers and abstracts addressing different aspects of generic questionnaires, the majority of the papers were descriptive. The evidence is not yet available to document that any of the generic instruments pass the requirements of the OMERACT Filter. (J Rheumatol 1999;26:210–6)

Key Indexing Terms:
METAANALYSES   QUALITY OF LIFE   RHEUMATIC DISEASES

Health related quality of life (HRQOL) instruments can be classified as either "disease-specific," where the comparisons are within a condition such as different therapies within rheumatoid arthritis (RA), or "generic," where comparisons can be made across different conditions in different disciplines, e.g., a therapy of RA versus a therapy in heart disease or cancer. During the OMERACT II conference, participants voiced concern over the validity of generic quality of life (QOL) measures in their current form, and suggested extensive validation in various musculoskeletal diseases before these measures could be recommended for decision making[1].

An OMERACT–International League of Associations for Rheumatology (ILAR) Task Force was created to study the available evidence, and make recommendations for the necessary studies. This report describes the implementation of a systematic review of all literature available that has used generic instruments alone or in comparison with disease-specific instruments, to assess the extent to which the requirements of the OMERACT filter are met[2].

MATERIALS AND METHODS

To identify studies that have used generic instruments as evaluative instruments in rheumatic diseases, alone or in comparison with disease-specific instruments, we performed MEDLINE and EMBASE searches from January 1988 to December 1997. Details of the search strategies are given in Appendix A. Studies were selected by one reviewer (ZO) and checked by a second (BS). Language was restricted to English articles.

The computer search was complemented by a hand search of the abstracts listed under "quality of life" at American College of Rheumatology (ACR) meetings over 5 years (1992 to 1997). After the papers were identified, the eligibility criteria were applied by two of the authors (BS, ZO). The criteria for including primary studies were: (a) randomized controlled trials (RCT) using generic instruments in populations older than 18 years with any of the following rheumatic diseases: RA, fibromyalgia (FM), osteoporosis, osteoarthritis (OA), systemic lupus erythematosus, and ankylosing spondylitis (AS); (b) observational or analytic studies that have compared generic versus specific instruments. Studies using only disease-specific instruments were excluded.

Three observers (BS, MGD, ZO) classified the papers as RCT, clinical trials, cohort studies, case-control, and case-series studies. Data were inde-

pendently extracted and cross-checked; disagreements were resolved by consensus.

RESULTS

The results of the searches are shown in Table 1. We retrieved 312 articles from MEDLINE. One hundred eighty-three articles were selected. Thirteen reports of 10 RCT that described the use of generic instruments[3-16] (Table 2) were analyzed. Sixty-six papers were retrieved from EMBASE, 26 papers not already identified on MEDLINE contained data on generic instruments. Of these, 6 papers described the use of generic instruments in RCT. These papers were also found in MEDLINE. There were 101 abstracts on quality of life in ACR abstract books; 78 abstracts contained data on generic instruments, and of these, 9 describe their use in RCT[17-25]. The characteristics of the RCT are described in Table 3.

From the set of 13 studies Thompson, et al[3] was not further considered, as it was part of a multicenter RCT published in 1986 by Bombardier, et al[4]. This study was a 6 month randomized trial at 14 sites, and the cost effectiveness of auranofin treatment for patients with RA was compared with placebo. Five endpoints were addressed: (a) traditional clinical rheumatologic measures; (b) nontraditional indicators of effect on specific aspects of RA; (c) effects in capability in activities of daily living (the instrument used to gauge effect on activities of daily living was the Health Assessment Questionnaire, HAQ); (d) assess-

ment of overall health status, measured with the Quality of Well Being Scale (QWB); and (e) economic measures of health care utilization. Descriptions and discussion of the endpoints, except the last one, were given in the Bombardier article (see below)[4].

Tugwell, et al[5] reported on the quality of life results in a double blind, randomized trial of methotrexate versus placebo in RA. The effect of treatment on physical, social, and emotional function was measured in 2 different ways: the same, standard measurements in all patients: the Lee Functional Index and the McMaster Health Index Questionnaire (MHIQ); and individualized measurements, i.e., MACTAR (McMaster-Toronto Arthritis Patient Function Preference) Questionnaire, selected by the patients at the start of the trial as representing the functions they most wanted to have improved by treatment. This set of measures addressed whether improvement had occurred in the leading physical disabilities that interfered most with activities of daily living (including mobility, self-care, work, and leisure activities). On the standard measurements, methotrexate treated patients scored better than placebo treated patients in their physical, social, and emotional function by 11, 5, and 6%, respectively, results that, although statistically significant, were small. However, in the MACTAR individualized measures methotrexate treated patients were 29% better, a result that was both highly statistically significant and greater than the differences in the standard measurements or in joint counts, grip strength, proximal interphalangeal joint circumference, morning stiffness, or walking time. Because the individualized measurements were as efficient as the best direct joint examination measures, yet reflected functional outcomes of greatest importance to individual patients, they constitute useful measures for such trials.

Bombardier, et al[6] compared instruments with respect to their responsiveness in detecting a treatment effect, the time involved in administering the instrument, the need for the presence of an interviewer, and each time of administration. Fourteen clinical centers in the United States and Canada participated in a 6 month, double-blind, placebo controlled study of auranofin as a treatment for patients with RA. Two hundred ninety-four patients were included in the study. Twenty-eight instruments were used to measure patients' pain and their clinical, functional, and global response to treatment. The measures of outcome included clinical function, pain, global impression, overall health, utility, and others, such as the National Institute of Mental Health Depression, Rand General, Health Perception, and erythrocyte sedimentation rate (ESR). The counts of tender and swollen joints were found to be the most responsive clinical measures, the 10 cm pain line was the most responsive and the fastest to administer of the pain instruments, and the categorical self-assessment of arthritis was the most responsive global measure. In the functional assessment category,

Table 1. Strategy searches results.

Source	Total Studies	Number of Selected Studies (%)
MEDLINE	312	183 (59)
EMBASE	176	66 (37)
ACR meeting abstracts	101	78 (77)
Total	413	387 (94)

Table 2. Results of MEDLINE search over the last 10 years, selection of studies, and number of RCT reports retrieved.

Publication Year	MEDLINE Articles	Selected Studies	RCT Reports
1988	16	12	1
1989	14	12	0
1990	4	4	1
1991	27	15	1
1992	31	21	1
1993	48	21	2
1994	17	18	4
1995	38	22	2
1996	84	28	1
1997	103	30	0
Total	312	183	13

Table 3. Principal characteristics of RCT evaluating QOL in rheumatic diseases.

Author	Patient Population	Study Design	Intervention and Followup (FU)	Outcome and Measurements	Results and Comments
Tugwell et al, 1990[5]	123 with RA	DB RCT	Methotrexate vs placebo FU 4.5 mo	Health status: Lee FI, MACTAR Patient Function Preference, MHIQ. Clinical: number of tender or swollen joints, pain, 50 ft walk, morning stiffness	MACTAR Patient Preference was clearly responsive to change
Bombardier et al, 1991[6]	294 with RA	DB RCT	Auranofin vs placebo FU 6 mo	Clinical: Number of tender or swollen joints, 50 ft walk. Health status: HAQ, Keitel Assessment, QWB, Toronto Activities of Living Questionnaire. Pain: McGill Pain Questionnaire, Pain Ladder Scale 10 cm VAS. Overall health: Arthritis Categorical Scale, Arthritis Ladder Scale, Overall Health Ladder Scale, Rand Current Health Assessment, 10 cm VAS by patient and physician, NIMH Depression Questionnaire, Rand General Health Perceptions Questionnaire	Several HRQOL measures were shown to be responsive change in RA due to therapy
Calfas et al, 1992[7]	40 with OA	RCT	Cognitive-behavioral modification vs traditional education intervention FU: 2–6–12 mo	Health status: QWB. Physical function, psychological status, pain: AIMS. Depression: BDI. Social support: SSQ	Non-statistical difference was seen at one year, maybe because of the small sample
Laupacis et al, 1993[8]	188 after hip replacement	DB RCT	Cement vs cementless prosthesis FU 3 to 24 mo	HRQOL: Harris hip score, Merle d'Aubigné hip score, WOMAC OA Index, MACTAR Patient Preference Disability Questionnaire, TTO Physical function: 6 min walk	Marked improvement in physical function, social interaction, and overall health after hip replacement with both types of prosthesis
Rorabeck, et al, 1994[9]	164 with OA	RCT	Cement vs cementless prosthesis FU 24 mo	HRQOL: Harris hip score, Merle d'Aubigné hip score, WOMAC OA Index, MACTAR Patient Preference Disability Questionnaire, SIP, TTO	No difference in improvement after 2 years
Hidding, et al, 1993[10]	135 with AS	RCT	Supervised group physical therapy 9 mo vs unsupervised Individualized FU 9 mo	Spinal mobility. Fitness by ergometry. Functioning: SIP, HAQ for AS, and FI. Global assessment: 10 cm VAS	Improvement for all primary outcomes, greater for mobility, fitness, and global health in group therapy. Difference statistically significant
Hiddings et al, 1994[11]	135 with AS	RCT	Cessation vs continue weekly group physical therapy FU 9 mo	Spinal mobility: 10 cm segment method, chest expansion, and goniometer measured cervical rotation. Physical fitness: ergometric. Functioning: SIP, HAQ-S, FIS. Global health: VAS. Pain and stiffness: VAS	Significant improvement only in global health and in SIP
Bakker, et al, 1994[12]	59 with AS	RCT	Weekly sessions of supervised group physical therapy for a period of 9 mo vs daily exercises at home FU 9 mo	HRQOL: MUMQ, SIP, HAQ-S, AIMS. Physical function: spinal mobility, enthesis and articular indices	Standard gamble utilities address different aspects of health status. Utility measurement is sensitive to the method chosen to elicit patient well being
Bakker, et al, 1995[13]	134 with AS and 73 with FM	RCT	134 AS patients were randomly allocated to weekly sessions of group physical therapy or daily exercises at home. 73 FM patients randomized into one of 3 groups: low impact fitness, biofeedback, and controls. FU 6 and 9 mo	Health status: MUMQ, SIP, HAQ-S, AIMS. Patient priorities: Problem Elicitation Technique (PET)	Obtaining patients' priorities (PET) was feasible, its construct validity was satisfactory, and it was more responsive to change in patients with AS than in patients with FM
Burckhardt, et al, 1994[14]	86 with FM	RCT	Education vs physical therapy FU 3 mo	HRQOL: FAI, QOLS Swedish SELF. Physical fitness: 6 min walk, chair test	Self-efficacy significantly enhanced. Other changes smaller than expected
Buchbinder, et al, 1995[15]	112 with AR	RCT	Low dose cyclosporine vs placebo FU 6 mo	Tender and swollen joint count. Morning stiffness. ERS. Pain: 10 cm VAS. Functional status: PET (MACTAR preference disability questionnaire modified), HAQ, AIMS	Evidence supports the core set of outcome measures proposed by ACR and OMERACT
Maggs, et al, 1996[16]	172 with chronic arthritis	RCT	Self-instruction with aid of educational booklet with or without one-to-one instruction FU 6 weeks	Health status: NHP. Functioning: HAQ (British adaptation). Arthritis knowledge: questionnaire	Significant increase in knowledge not associated with improved clinical status. No change in QOL measurements. One-to-one instruction showed no differences.

AIMS: Arthritis Impact Measurement Scales; AS: Ankylosing Spondylitis; BDI: Beck Depression Inventory; FAI: Fibromyalgia Attitudes Index; FI: Functional Index; HAQ: Health Assessment Questionnaire; MACL: Mood Adjective Check List; MACTAR: McMaster-Toronto Arthritis; MHIQ: McMaster Health Index Questionnaire; NIMH: National Institute of Mental Health; NHP: Nottingham Health Profile; OA: Osteoarthritis; OMERACT: Outcome Measures in Clinical Trials; TTO: Time Trade Off; QOL: Quality of Life; QOLS: Quality of Life Scale; QWB: Quality of well being; SELF: Self-Efficiency Scale; SIP: Sickness Impact Profile; SSQ: Social Support Questionnaire; VAS: visual analog scale; WOMAC: Western Ontario and McMaster University.

the Health Assessment Questionnaire (HAQ), the Keitel Functional Assessment, and the QWB Questionnaire were equally responsive. The HAQ was the shortest and the only self-administered questionnaire. The QWB has had the most extensive validation work, but is a time consuming instrument, requiring intensive interviewer training. The Keitel was another time consuming instrument, but had the advantage of high interobserver agreement. The results indicate improvement in the number of inflamed joints is accompanied by meaningful improvements across a range of outcomes relevant to the patient's quality of life. It is also noted that these improvements can be detected as efficiently as can those in the clinical measures. The overall result was that several HRQOL measures were shown to be responsive to change in RA due to therapy.

Calfas, et al[7] compared a cognitive-behavior modification and a traditional education intervention for adults with OA. Forty patients with OA were randomly assigned to one of two groups: cognitive-behavior modification or didactic lectures. During 10 weekly sessions, the cognitive-behavior group learned methods for coping with pain and the disabilities associated with OA. The traditional education group were given a series of lectures from health care professionals. Prior to the interventions and following 2, 6, and 12 months, patients in both groups were evaluated with a general QWB scale, the Arthritis Impact Measurement Scale (AIMS), the Beck Depression Inventory (BDI), and other measures. Although there were some differences between the 2 groups at the 2 month followup, by the end of one year, physical and psychological functioning did not differ significantly between the 2 groups. In comparison to baseline, both groups demonstrated changes at 2 months on QWB, depression, and the pain component of the AIMS. Improvements in depression remained through the one year followup. Multiple regression analysis revealed that the mobility and physical activity aspects of the AIMS were significant longterm predictors of outcome (one year) for general QOL measures. One year outcomes for depression were significantly predicted from scores on social support and mobility measures from the AIMS. The conclusion was that cognitive-behavior modification and education produce similar effects on longterm physical and psychological functioning in patients with OA. The results obtained at 12 months with the QWB instrument were not significant; however, the authors did mention the small sample size as a limitation.

Laupacis, et al[8] examined the effect of 2 types of prosthesis on the health related QOL of patients who have OA. One hundred eighty-eight patients were followed for 3 months, 179 for 6 months, 156 for one year, and 90 for 2 years. The HRQOL was assessed with use of the Harris hip score, the Merle d'Aubigné hip score, the Sickness Impact Profile (SIP), the Western Ontario and McMaster University (WOMAC) Osteoarthritis Index, the MACTAR Question-

naire, and the time tradeoff technique (TTO) as a measure of utility. Patients also took the 6 min walk test. There was significant improvement in all HRQOL measures and in the 6 min walk test after the operation [p < 0.01 for all items, except for the work dimension of the SIP at 3 months (p = 0.07)]. Most of the improvement had occurred by 3 months postoperatively. At 2 years, the mean distance walked in the 6 min walk test had increased from 247 to 408 meters; the mean Harris score had improved from 44 to 98 points; the mean Merle d'Aubigné score from 10 to 17.7 points; the mean global physical score of the SIP from 23.1 to 3.2 points; the mean pain score of the WOMAC from 4.9 to 0.7 points; the mean MACTAR from 7.7 to 0.8 points; and the mean TTO score from 0.32 to 0.87. The responsiveness of generic instruments correlates with those of disease-specific instruments. This study confirms the marked improvement in physical function, social interaction, and overall health that occurs after hip replacement, as well as the feasibility of the performance of randomized trials to compare 2 types of prostheses. Two year followup results show maintenance of improvement[9].

Hidding, et al[10] carried out a RCT to study the effects of adding supervised group physical therapy to unsupervised individualized therapy in ankylosing spondylitis. This study was followed by another RCT conducted by Hidding[11] in which the aim was to determine whether beneficial effects persisted after cessation of the group physical therapy intervention. Before randomization, all patients received supervised individual therapy for 6 weeks to standardize their pretreatment condition and to better reflect clinical practice. One hundred forty-four patients were randomized to exercises at home, or the same plus weekly group physical therapy for 9 months. One blinded observer at baseline and subsequently every 3 months up to 9 months assessed all patients. Endpoints were spinal mobility, fitness (maximum work capacity by ergometry), functioning [SIP, HAQ for ankylosing spondylitis (HAQ-S), Functional Index], and patient global assessment of change on a 10 cm visual analog scale. The results showed that spinal mobility improved. Global assessment improved by 0.3 (6%) after home exercises, and by 1.7 (34%) after group therapy. There were no significant differences in chest expansion, cervical rotation, or self-assessment of functioning. SIP and HAQ-S showed baseline scores of zero in one-fourth of the patient population. Although these patients were excluded from the analyses no significant effect was found. These results raise questions about the utility of SIP and HAQ-S for clinically relevant improvement; however, sample size and other characteristics of the study may be the answers. After a 9 month period of supervised group physical therapy, 68 patients were randomized for another 9 months to unsupervised daily exercises at home (discontinuation group) or continuation of weekly sessions of supervised group physical therapy (continuation group). Endpoints were the same as

those used in the first report[10]. The results showed that time for exercises at home was significantly higher in the continuation than in the discontinuation group (mean duration 1.9 versus 1.2 h per week, p < 0.05). The continuation group improved in global health (mean improvement 1.6, 32%) and in SIP score. For both studies combined, the total mean improvement showed a decrease in dysfunctioning from 2.6 to 1.2 in SIP score. Scores for thoracolumbar mobility and HAQ-S did not change very much, whereas chest expansion, cervical rotation, fitness, and FI deteriorated. The average attendance for group therapy sessions was 62%, lower than the average attendance in group therapy in the first 9 month period (74%). The discontinuation group improved only marginally (0.2) (4%) in global health, whereas all other endpoints decreased. Only for global health and HAQ-S were the differences statistically significant in favor of the continuation group. The authors stated that global health and functioning are sustained or even improved further if group physical therapy is continued.

Bakker, et al[12] published the results of a study that used patients included in the Hidding study[10]. The aim was to compare utilities derived by a rating scale and the standard gamble method in patients with ankylosing spondylitis, to relate these values to other outcome measures, and to assess the responsiveness to change of utilities relative to changes in other outcomes. Analysis was restricted to the 59 patients who completed the Maastricht Utility Measurement Questionnaire at baseline and after 9 months. Reliability was assessed by intraclass correlation coefficient and change scores for marker states of disease. Construct validity was evaluated by correlation and multiple regression of baseline values with a variety of disease outcomes (pain and stiffness, patient and physician global assessment, SIP, HAQ-S, AIMS, functional articular, and enthesis indices and spinal mobility measures). The test–retest intraclass correlation coefficients for patient utilities were 0.95 (rating scale) and 0.79 (standard gamble), and for the marker state of mild disease 0.70 (rating scale) and 0.77 (standard gamble). A multiple regression analysis with the baseline rating scale or standard gamble utilities as dependent variable showed that patient's global assessment explained 59 and 11% of the total variance, respectively. By multiple regression analysis 10% of the variance of change in rating scale utilities was explained by changes of patient global assessment. In contrast, variance in change in standard gamble utilities was not explained by changes in other disease outcomes. Findings obtained by rating scale and standard gamble differ considerably. Standard gamble utilities seem to address different aspects of health status than do rating scale utilities and more traditional outcomes. Utility measurement is sensitive to the method chosen to elicit patient well being.

In 1995, Bakker, et al[13] published the results of a Problem Elicitation Technique (PET) study. The sample population was obtained from the Hidding article[10], as well as 73 patients with FM. Trained interviewers at baseline and at 6 and 9 month followup, following physical therapy, applied the PET questionnaire. A PET score was calculated at each assessment; the maximum score was 49 and the minimum zero. A high PET score indicates a higher degree of perceived disability. Construct validity of the PET was assessed by correlation and multiple regression of baseline values with other disease outcomes (pain, stiffness, patient global assessment, SIP, HAQ, AIMS, patient utilities). Responsiveness to change of PET was assessed against changes in these outcomes and by comparing the efficiency of the PET with other outcomes. Patients with FM identified more problems (mean 6.8) than patients with AS (mean 4.4). The PET score improved from 14.9 to 11.3 (p = 0.0001) in patients with AS, but did not change [21.8 to 21.1 (p = 0.24)] in patients with FM. Construct validity testing of the PET score showed statistically significant (p < 0.05) correlation with AIMS, utilities, SIP, HAQ, pain, stiffness, and patient global health in both groups of patients (r varying from 0.22 to 0.66). Assessment of responsiveness to change revealed that PET score efficiency, defined as the mean change in the measure divided by the standard deviation of the change measure, was 0.6 in patients with AS and 0.09 in those with FM. These results indicated that the PET score was nonresponsive in patients with FM. However, all other measures were nonresponsive in these patients. In patients with AS spinal mobility, physical fitness, and patient assessed improvement were the most responsive. It remains unclear if the lack of responsiveness in FM was due to inefficiency of therapy or nonresponsiveness of the instruments.

Burckhardt, et al[14] conducted a RCT to determine the effectiveness of self-management education and physical training in decreasing FM symptoms and increasing physical and psychological well being. A pretest–post test control group design was used. Ninety-nine women with FM were randomly assigned to one of 3 groups; 86 completed the study. The education-only group received a 6 week self-management course. The education plus physical training group received the course and 6 h of training designed to prepare them to exercise independently. The control group received treatment after 3 months. A set of instruments to measure QOL was used including the Fibromyalgia Impact Questionnaire (FIQ); Fibromyalgia Attitudes Index (FAI); Quality of Life Scale (QOL-S), Self-Efficacy Scale (SELF); and Beck Depression Inventory (BDI). The results showed that experimental programs had a significantly positive effect on QOL and self-efficacy. Helplessness, number of days feeling bad, physical dysfunction, and pain in the tender points decreased significantly in one or both of the treated groups when retested 6 weeks after the end of the program. A longterm followup of 67 treated subjects showed significant positive changes on the FIQ physical function in the physical training group. Scores on SELF and QOL-S

were significantly higher on followup in the group that got both education and physical training. These results correlated with the FAI score. Other changes were smaller and more delayed than had been expected. The authors recognized several limitations in this study and suggested that future trials should include a longer education program, more vigorous physical training, and longterm followup.

Buchbinder, et al[15], utilizing data from a multicenter randomized double blind clinical trial of low dose cyclosporine and placebo in 144 patients with severe RA, estimated the relative efficiency of measures to detect treatment effect. Four pain measures (10 cm VAS, 5 point categorical scale, HAQ pain index, AIMS pain score) and 3 QOL measures (PET, HAQ, AIMS) were compared. Physician and patient global measures were the most responsive instruments, although neither was statistically superior to tender joint count. Swollen joint count, grip strength, pain measured on a 10 cm VAS, and functional status measured by the PET and HAQ were all of intermediate responsiveness. Morning stiffness, 5 point pain scale, and ESR were the least responsive instruments.

Maggs, et al[16] described the effect on knowledge and health status of giving patients a booklet on arthritis with or without instruction by health professional. Consecutive new attendees at a rheumatology clinic were randomly allocated to one of 3 groups. All groups received routine care, but one received no other intervention, one an educational booklet on arthritis, and one the booklet plus instruction from a health professional. Prior to intervention, all groups had similar knowledge, Nottingham Health Profile (NHP), and HAQ score. After 6 weeks, the knowledge score was significantly increased in both groups given the booklet, but not in the control group. The group instructed by a health professional showed no greater increase than the group given the booklet alone. Increased knowledge was not associated with improved clinical status and no group showed a significant change in NHP or HAQ scores. The time period was short and may have allowed insufficient time for changes in knowledge to affect outcome measures. Nearly all patients said they found the booklet useful.

DISCUSSION

The results derived from the use of generic instruments in clinical trials are important data needed by clinicians and policy makers to understand the tradeoff in resource allocation. Systematic reviews and metaanalyses, such as those contained in the Cochrane Library, increasingly contain data on generic QOL instruments[26]. It is important to show that generic instruments are responsive to change in musculoskeletal disease so that inappropriate policy decisions allocating resources to other areas are not made. Governmental agencies, such as the US Food and Drug Administration, are now proposing to accept distinct claims of the effect of drugs and biologic agents upon QOL. This would require

evidence of a clinically important improvement upon the generic QOL scale. The core set of OMERACT outcomes in RA may have to expand to include generic QOL instruments. However, since these are required to be evidence based, adequate documentation of their ability to meet the requirement of the OMERACT Filter[2] is a prerequisite.

Although there has been a substantial increase in the number of articles and abstracts addressing different aspects of generic questionnaires, the majority of these were based on description rather than intervention studies. The current data from trials are promising in that clinically meaningful change in disease-specific measures were often reflected in smaller, but nonetheless statistically significant changes in some generic measures. Developing a coherent conceptual framework, and evaluating responsiveness or clinical importance of a change score, should be a major priority if we are to keep up with other disease areas such as cardiovascular disease and cancer. Therefore, we encourage people to collect the data required for this documentation by including at least one generic instrument in all new Phase III and IV controlled trials.

APPENDIX A

(Rheumatoid Arthritis) or (explode "Arthritis,-Rheumatoid"/all subheadings)
(Osteoarthritis) or (explode "Osteoarthritis"/all subheadings)
(Fibromyalgia) or (explode "Fibromyalgia"/all subheadings)
(Osteoporosis) or (explode "Osteoporosis"/all subheadings)
(explode "Rheumatic-Diseases"/all subheadings) or Rheumatic diseases
("Spondylitis,-Ankylosing"/all subheadings) or (Ankylosing Spondylitis) or
(Spondyloarthropathy)
(exact{SF-36}) or (Short Form Health Survey)
(Sickness Impact Profile) or ("Sickness-Impact-Profile")
Health Utility Index
Nottingham Health Profile
Feeling Thermometer
Standard Gamble
Time Trade-Off
explode "Quality-of-Life"/all subheadings

REFERENCES

1. Tugwell P, Boers M, Brooks PM. OMERACT II Conference — outcome measures in RA clinical trials: conclusion. J Rheumatol 1995;22:1431–2.
2. Boers M, Brooks P, Strand CV, Tugwell P. The OMERACT filter for outcome measures in rheumatology. J Rheumatol 1998;25:198–9.
3. Thompson MS, Read JL, Hutchings HC, Paterson M, Harris ED Jr. The cost effectiveness of auranofin: results of a randomized clinical trial. J Rheumatol 1998;15:35–42.
4. Bombardier C, Ware J, Russell IJ, et al. Auranofin therapy and quality of life in patients with rheumatoid arthritis: results of a multicenter trial. Am J Med 1986;81:565–78.
5. Tugwell P, Bombardier C, Buchanan WW, et al. Methotrexate in rheumatoid arthritis. Impact on quality of life assessed by traditional standard-item and individualized patient preference health status questionnaires. Arch Intern Med 1990;50:59–62.
6. Bombardier C, Rabat J, and Auranofin Cooperating Group. A comparison of health related quality of life measures for rheumatoid arthritis research. Control Clin Trials 1991;12:243S–256S.

7. Calfas K, Kaplan R, Ingram R. One-year evaluation of cognitive-behavioral intervention in osteoarthritis. Arthritis Care Res 1992;5:202–9.

8. Laupacis A, Bourne R, Rorabeck C, et al. The effect of elective total hip replacement on health-related quality of life. J Bone Joint Surg Am 1993;75:1619–26.

9. Rorabeck CH, Bourne RB, Laupacis A, et al. A double-blind study of 250 cases comparing cemented with cementless total hip replacement. Cost-effectiveness and its impact on health-related quality of life. Clin Orthop Rel Res 1994;298:156–64.

10. Hidding A, van der Linden S. Is group physical therapy superior to individualized therapy in ankylosing spondylitis? A randomized controlled trial. Arthritis Care Res 1993;6:117–25.

11. Hidding A, van der Linden S, Gielen X, de Witte L, Dijkmans B, Moolenburgh D. Continuation of group physical therapy is necessary in ankylosing spondylitis: results of a randomized controlled trial. Arthritis Care Res 1994;7:90–6.

12. Bakker C, Rutten van Molken M, Hidding A, van Doorslaer E, Bennett K, van der Linden S. Patient utilities in ankylosing spondylitis and the association with other outcome measures. J Rheumatol 1994;21:1298–304.

13. Bakker C, van der Linden S, van Santen Hoeufft M, Bolwijn P, Hidding A. Problem elicitation to assess patient priorities in ankylosing spondylitis and fibromyalgia. J Rheumatol 1995;22:1304–10.

14. Burckhardt CS, Mannerkorpi K, Hedenberg L, Bjelle A. A randomized, controlled clinical trial of education and physical training for women with fibromyalgia. J Rheumatol 1994;21:714–20.

15. Buchbinder R, Bombardier C, Yeung M, Tugwell P. Which outcome measures should be used in rheumatoid arthritis clinical trials? Clinical and quality-of-life measures' responsiveness to treatment in a randomized controlled trial. Arthritis Rheum 1995;38:1568–80.

16. Maggs FM, Jubb RW, Kemm JR. Single-blind randomized controlled trial of an educational booklet for patients with chronic arthritis. Br J Rheumatol 1996;35:775–7.

17. Zhao S, Hatoum H, Hubbard R, et al. Effect of celecoxib, a novel Cox-2 inhibitor, on health-related quality of life of patients with osteoarthritis of the knee [abstract]. Arthritis Rheum 1997;40 Suppl:S88.

18. Villanueva I, Navarro F, Toyos FJ, Ariza R, Guzman MM. Validation of a Spanish version of the WOMAC (Western Ontario and McMaster Universities Osteoarthritis Index) [abstract]. Arthritis Rheum 1997;40 Suppl:S110.

19. Carlin B, Noe L, Lookabaugh J, et al. Quality of life in active rheumatoid arthritis patients receiving recombinant human interleukin-1 receptor antagonist [abstract]. Arthritis Rheum 1994;37 Suppl:S249.

20. Peeters H, Jongen-Lavrencic M. Recombinant human erythropoietin improves health related quality of life in patients with rheumatoid arthritis and anemia of chronic disease: utility measures correlate strongly with disease activity measures [abstract]. Arthritis Rheum 1996;39 Suppl:S246.

21. Gazarian M, Ibanez D, Bombardier C. Relationship between cyclosporin adverse effects and quality of life measures in patients with rheumatoid arthritis [abstract]. Arthritis Rheum 1996;39 Suppl:S105.

22. Hatoum H, Ware J, Keller S, Kong S, Kuss M, Bocanegra T. Effect of Oxaprozin and Nabumetone on health-related quality of life of patients with osteoarthritis of the knee [abstract]. Arthritis Rheum 1996;39 Suppl:S225.

23. Carlson A, Morris L, Perfetto E, Sanda M, Vavra I, Lynch J. Quality of sleep for osteoarthritis patients using pharmaceuticals [abstract]. Arthritis Rheum 1996;39 Suppl:S262.

24. Gowans S, Voss S, deHueck A, Richardson M. A randomized controlled trial of exercise and education in fibromyalgia [abstract]. Arthritis Rheum 1996;39 Suppl:S91.

25. Tsakonas E, Fitzgerald A, Fitzcharles MA, et al. Results of a prospective long-term study of the effects of early institution of 2nd-line agents in rheumatoid arthritis [abstract]. Arthritis Rheum 1997;40 Suppl: S54.

26. The Cochrane Library [database on disk and CDROM]. The Cochrane Collaboration; Issue 1. Oxford: Update Software; 1998. Updated quarterly.