Generic Quality-of-life Assessment in Rheumatoid Arthritis

Peter Tugwell, MD, MSc; Leanne Idzerda, BAH; George A. Wells, PhD, MSc

Abstract

Background: Generic health status measures are commonly used in the evaluation of rheumatoid arthritis (RA) patients. The reliability, validity, and sensitivity of the instruments in the assessment of quality of life (QOL) in RA, and how they correlate to other clinical measurements, have long been questioned.

Objective: Analyze the performance of a commonly used generic health status measure, the Medical Outcomes Study 36-Item Short Form (SF-36), against the Outcome Measures in Rheumatology (OMERACT) criteria.

Methods: Data were analyzed from 7 double-blind, randomized controlled trials that examined the effectiveness of 1 or more interventions in RA. The primary outcome measures evaluated were the Mental and Physical Component Scores of the SF-36. Comparators were 1 or more of the following: the Health Assessment Questionnaire scores, tender joint count (TJC), the Disease Activity Score, and the American College of Rheumatology Responder Index (ACR20, ACR50, ACR70). The ability to detect a treatment effect in the study outcomes was evaluated using 3 measures: treatment difference, standardized response mean, and relative efficiency in relation to the TJC.

Results: As a generic QOL measure, the SF-36 is better suited to capture the holistic health of the patient, as reflected in the World Health Organization definition of health as being not only the avoidance of disease but the physical, emotional, and social well-being of the patient. Furthermore, use of the SF-36 permits comparisons of physical and mental aspects of QOL in the RA patient population, as well as comparisons of QOL parameters between patients with RA, other patient groups, and the general population.

Conclusion: The SF-36 deserves serious consideration for inclusion in the core set of outcomes in RA trials.

(Am J Manag Care. 2007;13:S224-S236)

Patient quality of life (QOL) outcome-based studies are designed to evaluate whether patient health has improved as measured by physical, mental, and social instruments.[1] In industrialized countries, only one third of the burden from disease is due to mortality, with two thirds due to physical, mental, and social disability.[2] Although the inverse is true in low- and middle-income countries, a third of the burden of disease in these domains is still due to the impact on well-being. Thus, we require appropriate outcome measures for those medical interventions that are designed to improve well-being in addition to, or instead of, extending the duration of the patient's life. This is certainly true for rheumatoid arthritis (RA) trials.

Of the many measurement tools available to clinical researchers, those that measure patient well-being are perhaps the most important for evaluating patient-perceived outcomes. Because well-being is a complex concept or attribute, its definition has been the subject of great debate.[3] It is variably interpreted as health-related QOL (HRQOL) or function. Fitzpatrick et al have distinguished "global definitions" from "component definitions" for patient-based outcome measures.[4] Global definitions define well-being in general terms such as global judgments of health or satisfaction with life, whereas component definitions break the concept into specific parts or dimensions. They have proposed the following classification of components: physical function, symptoms, psychologic well-being, social well-being, cognitive functioning, role activities, personal constructs (ie, life satisfaction, spirituality, etc), and satisfaction with care.

Measuring well-being poses challenges that are not apparent with more objective clinical outcome measurements. One of these challenges is measuring differences between individuals at a single point in time versus changes within individuals or groups over time. A discriminative instrument asks questions such as: who, at this point, has better QOL and whose QOL is not so good versus an evaluative instrument which asks who has improved more, who has improved less, and who has deteriorated. The latter is most frequently used in clinical trials.

The focus of this article is to classify HRQOL scales as disease-specific versus generic. Disease-specific scales are designed to be used for a spe-

Address correspondence to: Peter Tugwell, MD, MSc, Director, Centre for Global Health, University of Ottawa, 501 Smyth Rd, Rm LM12, Ottawa, ON, K1H8L6, Canada. E-mail: tugwell@uottawa.ca.

cific condition, with a scale for arthritis having different questions than a scale for heart failure. Disease-specific scales have typically been designed to identify aspects of a disease most likely to improve with therapy and thus will maximize a patient's responsiveness to change while receiving a particular therapy. All pivotal clinical trials of therapies now include an instrument to assess patient-reported outcomes, usually the Health Assessment Questionnaire (HAQ) or the Modified HAQ (MHAQ). Generic scales are designed to be applicable across many conditions and focuses on overall QOL (ie, overall physical, social, and emotional health). Because these are not tailored to a specific disease, generic scales are much less likely to be responsive to change in intervention trials. However, a number of trials of RA have included a generic scale, the Medical Outcomes Study 36-Item Short Form (SF-36).

Measurement Used in the Assessment of Quality of Life in Rheumatoid Arthritis Trials

Table 1 shows the most commonly used HRQOL instruments in RA trials and their psychometric properties.[5]

The HAQ and the MHAQ are the most widely used.[6] The HAQ Disability Index (HAQ-DI) is an ordinal scale with 20 items on daily functioning during the past week. These cover 8 component areas: dressing and grooming, arising, eating, walking, hygiene, reach, grip, and outdoor activities. The scale is either self-administered or may be applied in a personal or telephone interview. It can be completed in 5 minutes. Each response is scored on a 4-point scale of ability: without any difficulty, with some difficulty, with much difficulty, and unable to do.

There are a number of generic scales[7]; the SF-36 and its derivatives dominate the field in the majority of clinical areas, including RA trials. The SF-36 was designed as a generic indicator of health status for use in population surveys and evaluative studies of health policy, and is only more recently being used to complement disease-specific measures in clinical trials.[8] The SF-36 has 36 questions that measure the following 8 dimensions: physical functioning, physical role limitations, bodily pain, social functioning, general mental health, social role limitations, vitality, and general health perceptions. The standard instrument uses a 4-week recall period but is often used with a 1-week recall period. It may be self-administered or used in personal or telephone interviews. It takes 5 to 10 minutes to complete. Two summary scores are calculated, one for the physical component summary (PCS) and one for the mental component summary (MCS) scores. The focus of this article is to review the performance of the SF-36 against the Outcomes Measures in Rheumatology (OMERACT) Filter of Validity, Feasibility, and Responsiveness to Change.

What Is OMERACT?

OMERACT is an international, informally organized network initiated in 1992 that aims to improve outcome measurement in rheumatology. Consensus conferences, which are chaired by an executive committee, are held every 2 years and rotate around the globe. Data-driven recommendations are prepared and updated by expert working groups at these consensus conferences. Recommendations include core sets of measures for most of the major rheumatologic conditions.[9]

What Does OMERACT Do?

OMERACT strives to improve end point outcome measurement through a data-driven, iterative consensus process. Agreement regarding the use of standardized end points in randomized controlled trials and longitudinal observational studies is extremely important. Their use facilitates comparisons of outcomes across studies to provide the best estimates of benefit and safety for therapeutic interventions across differing patient populations.[10]

With the goal of improving outcome measurement, OMERACT organizes consensus conferences. The key characteristics of OMERACT conferences include a commitment to the data-driven interactive development of a majority alignment across relevant stakeholder groups on determining relevant health outcome domains and endorsing valid, responsive, feasible health outcome measures/scales in patients with musculoskeletal conditions.[11]

To reach a consensus over what should be measured and how (ie, what measures are applicable in trials for each clinical indication), OMERACT has developed the following procedure. First, the organizing committee polls experts and opinion leaders to generate interest in the topic at hand. These individuals then form a committee to guide the subsequent process. From the general domains of health status, defined by the *Ds* (*discomfort*, *disability*, *dollar cost*, *death*), specific domains are formulated for the topic in question. In each domain, measures are collected and tested for their applicability. The domains and the applicable measures form the basis for the consensus guidelines.

Currently, an initiative starts as a special interest group. A small group of experts initiates the research agenda with literature reviews and validation studies. At the conference, in informal discussions, the research agenda is prioritized and tasks are distributed among interested parties. The next step is a workshop, where studies are presented to help the formulation and selection of the domains. Again, agreement is reached on priorities in research to be performed. The final step is the module in which evidence (both from literature

Table 1. Characteristics of Commonly Used Instruments to Measure QOL in RA Patients

Questionnaire	No. of Items	Subscales	Scaling Method	Time for Scoring (min)	Time for Administration (min)	Health Dimension			Psychometric Properties				
						Psychologic	Social Function	Physical Function	Test-retest Reliability	Internal Reliability	Sensitivity Reliability	Convergent Validity	Discriminant Validity
AIMS	67	Mobility, physical activity, dexterity, social role, social activity, ADLs, pain, anxiety, depression	5-point Likert scale	<5	15	•	•	•	•	•	•	•	•
AIMS2	57	AIMS subscales + arm function work, social support	5-point Likert scale	<5	23								
MACTAR	14	Physical disability	Yes/No ranking	<5	10								
MHAQ	8	Physical disability	4-point Likert scale	<2	2								
HAQ-DI	20	Physical disability	4-point Likert scale	<5	5			•	•	•	•	•	•

QOL indicates quality of life; RA, rheumatoid arthritis; AIMS, Arthritis Impact Measurement Scale; ADLs, activities of daily living; MACTAR, McMaster Toronto Arthritis Patient Preference Disability Questionnaire; MHAQ, Modified Health Assessment Questionnaire; HAQ-DI, HAQ Disability Index.

and targeted studies) is presented and final selection of measures can take place.

Both in workshops and in modules, plenary presentations are complemented by small group sessions where participants express their views and preferences. These views are brought back to the plenary session, where a final consensus is formulated with the help of interactive voting using electronic touchpads. In modules, consensus implies agreement on domains or measures, and in workshops consensus means the formulation of a research agenda in areas where data-driven decisions cannot be made. The process is iterative, in that guidelines are forever preliminary, based on the assumption that future data (sometimes a direct result of the research agenda) will serve to refine or modify them. The work needed to justify a module with voting can be fast-tracked and achieved within 12 months if there are sufficient existing data on the performance of the instruments measuring the selected attributes. The new staging of starting with special-interest groups, with criteria for moving to a workshop and the additional requirements to warrant a module, all reflect

the expectation that the process can take up to 6 years or more. This has been the case with outcomes for adverse effects, which has been a focus at every OMERACT meeting since the second OMERACT in 1994.

The OMERACT Filter

When Is a Measure "Applicable"? A measure is considered "applicable" when it passes the OMERACT filter in its intended setting. The OMERACT filter can easily be summarized in only 3 words: truth, discrimination, and feasibility.[12] Each word represents a question to be answered of the measure, in each of its intended settings.

Truth. Is the measure truthful? Does it measure what it intends to measure? Is the result unbiased and relevant? This criterion captures the issues of face, content, construct, and criterion validity.

Discrimination. Does the measure discriminate between 2 situations that are of interest? The situations can be states at one time (for classification or prognosis) or states at different times (to measure change). This criterion captures the issues of reliability and sensitivity to change.

Feasibility. Can the measure be applied easily, given constraints of time, money, and interpretability? This criterion addresses the pragmatic reality of the use of the measure, one that may be decisive in determining a measure's success.

The SF-36 has already been demonstrated to meet the first and third criteria above. It has been extensively validated. Face validity is greatest for the component scores, which are much easier to interpret and use for clinical decision making than the 8 subscales. Content, construct, and criterion validity have been demonstrated, and it has been shown to meet the feasibility criteria in many trials in many conditions.[7] The outstanding challenge that this article addresses is to demonstrate its discrimination in RA through showing its responsiveness to change in trials against placebo.

SF-36 Comparison With Other Scales

Methods

Inclusion Criteria. For inclusion in this analysis, the identified studies had to be double-blind, randomized controlled trials that examined the effectiveness of 1 or more interventions targeting patients with RA. Studies needed to include as reported outcomes the SF-36 with their component scores, the HAQ, and tender joint count (TJC).

Exclusion Criteria. Studies were excluded when a control group was absent and when studies reported data on the same, or highly related, sample. In these cases (Table 2)[13-39] the study with the most complete data on the SF-36, HAQ, and TJC was reported. Trials were also excluded when the published report did not contain adequate data and the data could not be obtained from the original authors.

Identification of Relevant Articles

A total of 7 studies were identified through 2 searches and consultation with an expert in the field. The first search on PubMed retrieved 93 studies using keywords *rheumatoid*

Table 2. Characteristics of Excluded Trials

First Author, Year	Reason Excluded
Bilberg, 2005[15]	Nonextractable data
Cohen, 2001[16]	Same sample as Strand, 1999[37]
Eichler, 2005[13]	Nonextractable data
Genovese, 2005[17]	Same sample as Westhovens, 2006[38]/Wells, 2007[39]
Harrison, 2005[18]	No control group
Helliwell, 1999[19]	Nonextractable data
Hewlett, 2005[20]	Nonextractable data
Kaplan, 2005[21]	Nonextractable data
Keystone, 2004[22]	Nonextractable data
Kosinski, 2002[23]	No control group
Martin, 2006[24]	Nonextractable data
Mathias, 2000[25]	Nonextractable data
Scott, 1999[26]	Same sample as Strand, 1999[37]
Scott, 2000[27]	No control group
Sköldstam, 2003[28]	Nonextractable data
Strand, 2004[29]	Same sample as Strand, 2005[31]
Strand, 2004[30]	Same sample as Strand, 1999[37]
Strand, 2005[31]	No control group
Tijhuis, 2003[32]	Nonextractable data
Torrance, 2004[33]	Nonextractable data
Tugwell, 2000[34]	Same sample as Strand, 1999[37]
Tuttleman, 1997[14]	Nonextractable data
Weinblatt, 2003[35]	Same sample population as Torrance, 2004[33]
Wolfe, 2005[36]	No control group

arthritis and *SF-36* from 1998 to the present. A second search with a wider catchment area was run in MEDLINE, the Cochrane CENTRAL, and EMBASE databases using the Ovid platform. This search was created using Medical Subject Headings (MeSH) of the National Library of Medicine for RA using textword searching to identify other variations. The search was then limited through a filter to identify randomized controlled trials.

Outcome Measures. The primary outcome measures evaluated by this review were the MCS and PCS of the SF-36. The scores can range from 0 to 100, with higher scores indicating better QOL. Comparators were one or more of the following: the HAQ scores, TJC, the disease activity score, and the American College of Rheumatology (ACR) Responder Index (ACR20, ACR50, and ACR70).

Statistical Methods. The ability to detect a treatment effect in the study outcomes was evaluated using 3 measures: treatment difference—difference between the mean change in the treatment group and mean change in the placebo group; standardized response mean (SRM)—ratio of the treatment difference to the pooled standard deviation (SD) of the mean change in scores; and relative efficiency (RE) in relation to the TJC—square of the ratio of the *t* statistics which corresponds to squaring the ratio of the SRM for the outcome to the SRM for the TJC. An RE >1 would imply that the outcome is more efficient than the TJC in detecting a treatment effect. m_a and s_a are the mean and SD, respectively, of the change in scores from baseline in the abatacept group, and n_a is the number of patients in this group. Similarly, m_c, s_c, n_c are the corresponding values for the control group.

1. Treatment difference: $m_a - m_c$
2. SRM: $(m_a - m_c) \Big/ \sqrt{\left(\frac{1}{n_a} + \frac{1}{n_c}\right) \frac{(n_a - 1)s_a^2 + (n_c - 1)s_c^2}{n_a + n_c - 2}}$
3. RE: (SRM of outcome)^2/(SRM of TJC)^2

Data Extraction

Data on the mean change in scores from baseline, the SD, and the sample size were extracted from the identified published studies using standardized data extraction sheets. When exact data were not available, graphs were used to extract the required data. If the SD for the change from baseline was not reported, the relevant *P* value was recorded and used to calculate the SD. In addition to these measures, demographic details and measures of severity were recorded at baseline. The data were extracted by 2 independent persons and reviewed by a third to establish reliability. In trials where data were not reported in a form that allowed extraction, the original investigators were contacted via an electronic letter for more information. Nine first authors were approached. Six responded and 1 provided extractable data.

Results

Thirty-five of the identified studies met the criteria for inclusion after the screening of the results for both the searches (Table 3).[37-46] Of these identified 35 studies, 5 studies were included after the criteria of exclusion was applied. These results were further supplemented by an expert in the field who has compiled a bibliography of all SF-36 studies, totaling more than 10 000 published studies to date, and 2 additional studies were identified for a total of 7.

Cohen et al, 2006. "Rituximab for Rheumatoid Arthritis Refractory to Anti-tumor Necrosis Factor Therapy."[40] The primary efficacy end point was response, determined using the ACR 20% improvement criteria (ACR20) at 24 weeks. Secondary end points included responses on the ACR50 and ACR70. All patients had active, long-standing RA. At week 24, significantly more rituximab-treated patients than patients receiving placebo demonstrated ACR20 (51% vs 18%), ACR50 (27% vs 5%), and ACR70 (12% vs 1%) responses and moderate-to-good European League Against Rheumatism responses (65% vs 22%). Rituximab-treated patients also had clinically meaningful improvements in fatigue, disability, and HRQOL (Functional Assessment of Chronic Illness Therapy–Fatigue, HAQ-DI, and SF-36, respectively). ACR20 response rates over time showed a statistically significant separation between rituximab treatment and placebo by week 8; ACR50 and ACR70 responses over time showed a statistically significant separation by week 12 and week 16 of treatment, respectively.

Kosinski et al, 2000. "Determining Minimally Important Changes in Generic and Disease-specific Health-related Quality of Life Questionnaires in Clinical Trials of Rheumatoid Arthritis."[41] Patients with RA were enrolled in 2 double-blind, placebo-controlled clinical trials and completed the SF-36 modified health survey and the HAQ-DI at baseline and 6-week follow-up assessments. In the first trial, patients were randomized to receive misoprostol, diclofenac sodium, or placebo. In the second trial, patients were randomized to receive 3 different dose levels of celecoxib, a cyclooxygenase inhibitor, and placebo.

Data on 5 RA severity measures were also collected at baseline and at 6 weeks (patient and physician global assessments,

Table 3. Characteristics of Included Trials

First Author, Year	Duration	Details of Patients (Mean Age, % Female, Avg Years with Disease, % RF Factor Positive)	Treatment Groups	N	Baseline Measures SF-36, MCS, PCS, HAQ, and TJC
Cohen, 2006[40]	24 weeks	53; 81; 12; 79	MTX + placebo	209	TJC (68 assessed) = 33.0 ± 15.6 HAQ-DI = 1.9 ± 0.5 SF-36 = NA
		52; 81; 12; 79	MTX + rituximab	311	TJC (68 assessed) = 33.9 ± 15.1 HAQ-DI = 1.9 ± 0.6 SF-36 = NA
Kosinski, 2000[41]	6 weeks		Placebo	116	PCS = 31 (SD = 8.9) MCS = 48.9 (SD = 11.3) HAQ = 1.3 (SD = 0.68) TJC* 1-17 (25%) = 35.9% 18-34 (26%-49%) = 39.1% 35-68 (50%+) = 25.0%
			Celecoxib or misoprostol or diclofenac sodium	440	PCS = 31 (SD = 8.9) MCS = 48.9 (SD = 11.3) HAQ = 1.3 (SD = 0.68) TJC* 1-17 (25%) = 35.9% 18-34 (26%-49%) = 39.1% 35-68 (50%+) = 25.0%
Kremer, 2002[42]	24 weeks	57; 81; 12.7; 87.6	MTX + placebo	133	PCS = NA MCS = NA HAQ-DI = 1.5 ± 0.58 TJC = 26.4 ± 12.59
		56; 76; 10.5; 78.6	MTX + leflunomide	130	PCS = NA MCS = NA HAQ-DI = 1.6 ± 0.62 TJC = 26.9 ± 13.04
Lipsky, 2000[43]/ Maini, 2004[44]	54 weeks	51; 80; 11; 77	MTX + placebo	88	PCS = 27 ± 8 MCS = 47 ± 12 HAQ = 1.7 ± 0.6 TJC = 31 ± 18
		54; 81; 10; 84	MTX + infliximab 3 mg/kg q8w	86	PCS = 27 ± 7 MCS = 46 ± 11 HAQ = 1.8 ± 0.6 TJC = 32 ± 18
		52; 77; 9; 80	MTX + infliximab 3 mg/kg q4w	86	PCS = 25 ± 8 MCS = 48 ± 12 HAQ = 1.7 ± 0.6 TJC = 31 ± 15
		54; 77; 11; 82	MTX + infliximab 10 mg/kg q8w	87	PCS = 26 ± 7 MCS = 48 ± 11 HAQ = 1.7 ± 0.6 TJC = 32 ± 16
		52; 73; 12; 82	MTX + infliximab 10 mg/kg q4w	81	PCS = 27 ± 8 MCS = 47 ± 11 HAQ-DI = 1.7 ± 0.6 TJC = 34 ± 16

Table 3. Characteristics of Included Trials *(Continued)*

First Author, Year	Duration	Details of Patients (Mean Age, % Female, Avg Years with Disease, % RF Factor Positive)	Treatment Groups	N	Baseline Measures SF-36, MCS, PCS, HAQ, and TJC
Strand, 1999[37]	52 weeks	55; 70; 6.9; 64.8	Placebo	101	PCS = 29 MCS = 48 HAQ-DI = 1.3 TJC = NA
		53; 76; 6.5; 59.4	MTX	162	PCS = 30 MCS = 49 HAQ-DI = 1.3 TJC = NA
		54; 73; 7.0; 64.8	Leflunomide	157	PCS = 30 MCS = 47 HAQ-DI = 1.3 TJC = NA
Wells, 2007[39]/ Westhovens, 2006[38]	6 months	53; 80; 21.2; 73.3	DMARDs + placebo	133	PCS = 27.8 (SD = 6.3) MCS = 42.9 (SD = 11.9) HAQ = 1.8 (SD = 0.6) TJC = 32.8 (SD = 13.4)
		54; 77.1; 11.4; 72.9	DMARDs + abatacept	258	PCS = 27.5 (SD = 6.9) MCS = 41.3 (SD = 12.4) HAQ = 1.8 (SD = 0.6) TJC = 31.3 (SD = 13.0)
Zhao, 2000[45]/ Simon, 1999[46]	12 weeks	54; 73; 11; NA	Placebo	100	TJC = NA HAQ-DI = 1.4 (SD = 8.6) PCS = 29.1 (SD = 0.66) MCS = 46.9 (SD = 10.8)
		54; 74; 11; NA	Celecoxib 100 mg bid	152	TJC = NA HAQ-DI = 1.4 (SD = 0.65) PCS = 29.7 (SD = 8.0) MCS = 47.6 (SD = 11.1)
		55; 73; 11; NA	Celecoxib 200 mg bid	156	TJC = NA HAQ-DI = 1.5 (SD = 0.73) PCS = 29.5 (SD = 7.9) MCS = 45.3 (SD = 12.3)
		54; 72; 10; NA	Celecoxib 400 mg bid	134	TJC = NA HAQ-DI = 1.4 (SD = 0.72) PCS = 29.5 (SD = 8.3) MCS = 47.5 (SD = 11.6)
		55; 72; 10; NA	Naproxen 500 mg bid	135	TJC = NA HAQ-DI = 1.5 (SD = 0.67) PCS = 29.9 (SD = 8.9) MCS = 46.2 (SD = 11.6)

*Percent of patients that experienced tenderness in ≤25% of joints, >25% to <50% of joints, or ≥50% of joints.
RF indicates rheumatoid factor; SF-36, Medical Outcomes Study 36-Item Short Form; MCS, mental component summary; PCS, physical component summary; HAQ, Health Assessment Questionnaire; TJC, tender joint count; MTX, methotrexate; HAQ-DI, HAQ Disability Index; NA, not applicable; SD, standard deviation; DMARDs, disease-modifying antirheumatic drugs.

joint swelling and tenderness counts, and global pain assessment). Comparison of changes in the SF-36 and HAQ-DI scores was made between groups of patients known to differ in the level of change on each RA severity measure. With few exceptions, changes in the SF-36 and HAQ-DI scores were not the same between patients who differed in the level of change on each RA severity measure. Changes in the SF-36 and HAQ-DI scores were more strongly related to changes in the patient and physician global assessments and patient pain assessment than to changes in the joint swelling and tenderness counts.

Kremer et al, 2002. "Concomitant Leflunomide Therapy in Patients With Active Rheumatoid Arthritis Despite Stable Doses of Methotrexate. A Randomized, Double-blind, Placebo-controlled Trial."[42] The primary efficacy variable in this trial was the rate of achievement of ACR20 at the end of the study. SF-36s were completed as an end point analysis. Secondary outcomes included ACR50 and ACR70 responder rates at week 24.

Mean improvements in individual components of the ACR response criteria were statistically significant for the leflunomide group compared with the placebo group (P <.001). The mean change from baseline in overall physical function, measured with the HAQ-DI, was 0.42 for the leflunomide group (29% improvement), nearly twice the minimum clinically important difference, and 0.09 for the placebo group (5% improvement).

Lipsky et al, 2000. "Infliximab and Methotrexate in the Treatment of Rheumatoid Arthritis."[43] Patients with active RA who were on methotrexate (MTX) therapy were randomized to receive either infliximab or placebo. Clinical response at week 54 was defined according to the ACR20. The percentages of patients with an ACR50 and ACR70 were also assessed. Scores on the physical component subscales of the SF-36 were more than 2 SD below the score for the general US population of persons without chronic conditions.

Therapy with infliximab plus MTX resulted in a sustained reduction in symptoms and signs of RA and increased the function of patients, as measured by the HAQ or by the SF-36.

Maini et al, 2004. "Sustained Improvement Over Two Years in Physical Function, Structural Damage, and Signs and Symptoms Among Patients with Rheumatoid Arthritis Treated With Infliximab and Methotrexate."[44] The overall objective of this study was to evaluate the efficacy and safety of repeated administration of infliximab plus MTX over a 2-year period in patients who previously experienced an incomplete response to MTX. Efficacy measures included HAQ (physical function), SF-36 (HRQOL), total radiographic scores (structural damage), and the ACR20 (signs and symptoms). Improvement was shown in HAQ scores and PCS scores for the infliximab plus MTX group compared with the MTX-only group. Stability in the SF-36 MCS was observed for the infliximab plus MTX group. Radiographic scores showed median changes from baseline to week 102 as follows: 4.25 MTX only versus 0.50 infliximab plus MTX. ACR20 responses at week 102 were 40% to 48% for the infliximab plus MTX groups versus 16% for the MTX-only group.

Strand et al, 1999. "Function and Health-related Quality of Life."[37] This study assessed the efficacy of leflunomide or MTX compared with placebo in improving function and HRQOL in patients with active RA, and examined correlations between response status using ACR criteria to document improvement. Clinically meaningful and statistically significant improvement in measures of function, HRQOL (MHAQ scores), PCS of the SF-36, and work productivity were seen during treatment with leflunomide in comparison with placebo. MTX administration resulted in significant improvements in comparison with placebo in the MHAQ scores and HAQ-DI. Results of this study show that the magnitude of improvement in 6 of the 8 SF-36 scales in the leflunomide group met or exceeded 10 points, with the exceptions of the general health profile, which worsened in the placebo population, and mental health, which reflected numerical, but not statistically significant, improvements in comparison with placebo.

Wells et al, 2007. "Responsiveness of Patient-reported Outcomes Including Fatigue, Sleep Quality, Activity Limitation, and Quality of Life Following Treatment With Abatacept in Patients With Rheumatoid Arthritis."[39] The Abatacept Trial in Treatment of Anti-TNF Inadequate Responders (ATTAIN) was a phase 3, multicenter, 6-month trial comparing abatacept (n = 258) with placebo (n = 133) on a background of disease-modifying antirheumatic drug therapy in RA patients who had an inadequate response to anti-tumor necrosis factor (TNF) therapy. Moderate-to-large SRMs (≥0.6) were observed for physician global assessment, HAQ, SF-36 PCS, SF-36 bodily pain, and fatigue. REs for physician global assessment, SF-36 bodily pain, pain intensity, HAQ, SF-36 PCS, fatigue, and patient global assessment were all more responsive than TJC. The SF-36 MCS, swollen joint count, activity limitation, sleep, and C-reactive protein were less responsive. SF-36 PCS improved more than MCS.

Table 4. Treatment Differences, Standardized Response Means, and Relative Efficiencies

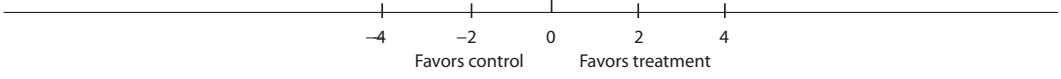
Outcome Study	Treatment Difference (95% CI)		Standardized Response Mean (95% CI)		Relative Efficiency
Physical Component Summary					
Cohen 2006[40]	4.90	(2.33 to 7.47)	0.33	(0.16 to 0.51)	0.23
Kosinski 2000[41]	–0.50	(–3.39 to 2.39)	–0.04	(–0.24 to 0.17)	0.02
Kremer 2002[42]	6.50	(2.65 to 10.35)	0.41	(0.16 to 0.65)	0.45
Lipsky 2000a[43]	1.89	(–0.36 to 4.14)	0.25	(–0.22 to 0.72)	0.28
Lipsky 2000b[43]	6.60	(4.21 to 8.99)	0.82	(0.34 to 1.30)	1.90
Lipsky 2000c[43]	7.60	(5.36 to 9.84)	1.01	(0.52 to 1.49)	2.55
Lipsky 2000d[43]	5.40	(2.97 to 7.83)	0.67	(0.19 to 1.15)	0.59
Strand 1999a[37]	6.60	(4.16 to 9.04)	0.68	(0.35 to 1.00)	1.35
Strand 1999b[37]	3.60	(1.10 to 6.10)	0.36	(0.04 to 0.68)	0.63
Wells 2007a[39]	3.98	(2.44 to 5.52)	0.42	(0.26 to 0.59)	0.70
Wells 2007b[39]	5.59	(3.66 to 7.51)	0.62	(0.40 to 0.83)	1.19
Zhao 2000a[45]	1.60	(–0.49 to 3.69)	0.19	(–0.23 to 0.62)	0.47
Zhao 2000b[45]	3.40	(1.34 to 5.46)	0.41	(–0.01 to 0.84)	1.49
Zhao 2000c[45]	3.50	(1.31 to 5.59)	0.41	(–0.02 to 0.84)	1.99
Zhao 2000d[45]	1.80	(–0.48 to 4.08)	0.20	(–0.22 to 0.63)	2.21
Mental Component Summary					
Cohen 2006[40]	3.40	(1.62 to 5.18)	0.33	(0.16 to 0.51)	0.23
Kosinski 2000[41]	1.70	(–1.36 to 4.76)	0.11	(–0.09 to 0.32)	0.26
Lipsky 2000a[43]	0.84	(–2.61 to 4.29)	0.07	(–0.40 to 0.54)	0.02
Lipsky 2000b[43]	1.04	(–2.55 to 4.63)	0.09	(–0.38 to 0.55)	0.02
Lipsky 2000c[43]	1.04	(–2.40 to 4.48)	0.09	(–0.38 to 0.56)	0.02
Lipsky 2000d[43]	0.94	(–2.57 to 4.45)	0.08	(–0.39 to 0.55)	0.01
Wells 2007a[39]	1.49	(–0.38 to 3.35)	0.13	(–0.03 to 0.29)	0.07
Wells 2007b[39]	3.71	(1.33 to 6.08)	0.33	(0.12 to 0.54)	0.34
Zhao 2000a[45]	2.50	(–0.28 to 5.28)	0.23	(–0.20 to 0.65)	0.65
Zhao 2000b[45]	3.50	(0.54 to 6.46)	0.30	(–0.13 to 0.72)	0.77
Zhao 2000c[45]	2.50	(–0.43 to 5.43)	0.22	(–0.21 to 0.65)	0.57
Zhao 2000d[45]	2.80	(–0.13 to 5.73)	0.25	(–0.18 to 0.68)	3.24
HAQ					
Cohen 2006[40]	–0.30	(–0.40 to –0.20)	–0.53	(–0.72 to 0.35)	0.58
Kosinski 2000[41]	–0.14	(–0.48 to 0.20)	–0.14	(–0.48 to 0.20)	0.40
Kremer 2002[42]	–0.33	(–0.48 to –0.21)	–0.68	(–0.93 to 0.43)	1.24
Lipsky 2000a[43]	–0.15	(–0.33 to 0.03)	–0.26	(–0.73 to 0.21)	0.29
Lipsky 2000b[43]	–0.34	(–0.52 to 0.16)	–0.56	(–1.04 to –0.09)	0.90
Lipsky 2000c[43]	–0.43	(–0.60 to –0.25)	–0.71	(–1.18 to –0.23)	1.25
Lipsky 2000d[43]	–0.39	(–0.57 to –0.21)	–0.65	(–1.13 to –0.17)	0.55
Strand 1999a[37]	–0.48	(–0.61 to –0.35)	–0.90	(–1.23 to –0.57)	2.37
Strand 1999b[37]	–0.29	(–0.42 to –0.16)	–0.54	(–0.89 to –0.22)	1.43
Wells 2007a[39]	–0.28	(–0.39 to –0.18)	–0.44	(–0.60 to –0.28)	0.77
Wells 2007b[39]	–0.33	(–0.44 to –0.22)	–0.64	(–0.85 to –0.42)	1.26
Zhao 2000a[45]	–0.07	(–0.17 to 0.03)	–0.13	(–0.42 to 0.16)	0.21
Zhao 2000b[45]	–0.19	(–0.30 to –0.08)	–0.31	(–0.60 to –0.02)	0.84
Zhao 2000c[45]	–0.18	(–0.29 to –0.07)	–0.30	(–0.59 to –0.01)	1.04
Zhao 2000d[45]	–0.12	(–0.22 to –0.02)	–0.22	(–0.51 to 0.07)	2.63
Tender Joint Count					
Cohen 2006[40]	–11.70	(–14.70 to –8.70)	–0.70	(–0.88 to –0.51)	—
Kosinski 2000[41]	–4.70	(–8.97 to –0.43)	–0.23	(–0.43 to –0.02)	—
Kremer 2002[42]	–7.60	(–10.61 to –4.59)	–0.61	(–0.86 to –0.36)	—
Lipsky 2000a[43]	–8.55	(–13.94 to –3.16)	–0.47	(–0.95 to 0.00)	—
Lipsky 2000b[43]	–9.92	(–14.88 to –4.96)	–0.60	(–1.07 to –0.12)	—
Lipsky 2000c[43]	–10.79	(–15.87 to –5.71)	–0.63	(–1.11 to –0.16)	—
Lipsky 2000d[43]	–14.97	(–20.16 to –9.78)	–0.87	(–1.36 to –0.39)	—
Strand 1999a[37]	–4.70	(–6.57 to –2.83)	–0.58	(–0.88 to –0.28)	—
Strand 1999b[37]	–3.60	(–5.45 to –1.75)	–0.45	(–0.75 to –0.16)	—
Wells 2007a[39]	–6.98	(–9.25 to –4.71)	–0.50	(–0.67 to –0.34)	—
Wells 2007b[39]	–8.06	(–11.06 to –5.05)	–0.57	(–0.78 to –0.35)	—
Zhao 2000a[45]	–4.00	(–6.57 to –1.43)	–0.28	(–0.57 to 0.01)	—
Zhao 2000b[45]	–4.80	(–7.37 to –2.23)	–0.34	(–0.63 to –0.05)	—
Zhao 2000c[45]	–4.10	(–6.69 to –1.51)	–0.29	(–0.58 to 0.00)	—
Zhao 2000d[45]	–1.90	(–4.43 to 0.63)	–0.14	(–0.43 to 0.15)	—

a, b, c, and d = treatment arms analyzed.
CI indicates confidence interval; HAQ, Health Assessment Questionnaire.

Figure 1a. Standardized Response Mean (SRM) of the SF-36 Physical Component Summary (PCS) Score

Review: SF-36
Comparison: 02 SRM
Outcome: 01 PCS

Study or subcategory	SRM (SE)	Weight (%)	SRM, random (95% CI)
Cohen 2006[37]	0.3345 (0.0900)	9.93	0.33 (0.16, 0.51)
Kosinski 2000[38]	–0.0355 (0.1044)	9.33	–0.04 (–0.24, 0.17)
Kremer 2002[39]	0.4092 (0.1246)	8.49	0.41 (0.16, 0.65)
Lipsky 2000a[40]	0.2501 (0.2395)	4.66	0.25 (–0.22, 0.72)
Lipsky 2000b[40]	0.8214 (0.2454)	4.52	0.82 (0.34, 1.30)
Lipsky 2000c[40]	1.0062 (0.2482)	4.46	1.01 (0.52, 1.49)
Lipsky 2000d[40]	0.6718 (0.2449)	4.53	0.67 (0.19, 1.15)
Strand 1999a[42]	0.6782 (0.1652)	6.89	0.68 (0.35, 1.00)
Strand 1999b[42]	0.3591 (0.1621)	7.01	0.36 (0.04, 0.68)
Wells 2007a[43]	0.4210 (0.0837)	10.19	0.42 (0.26, 0.59)
Wells 2007b[43]	0.6172 (0.1106)	9.07	0.62 (0.40, 0.83)
Zhao 2000a[45]	0.1935 (0.2161)	5.27	0.19 (–0.23, 0.62)
Zhao 2000b[45]	0.4144 (0.2165)	5.25	0.41 (–0.01, 0.84)
Zhao 2000c[45]	0.4139 (0.2191)	5.18	0.41 (–0.02, 0.84)
Zhao 2000d[45]	0.2045 (0.2180)	5.21	0.20 (–0.22, 0.63)
Total (95% CI)		100.00	0.42 (0.29, 0.55)



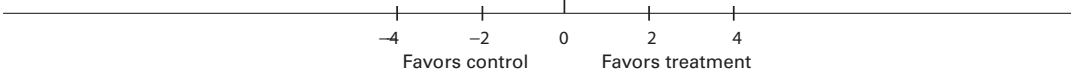
Test for heterogeneity: $Chi^2 = 36.95$, df = 14 ($P = .0008$), $I^2 = 62.1\%$.
Test for overall effect: Z = 6.44 ($P < .00001$).

a, b, c, and d = treatment arms analyzed.
SF-36 indicates Medical Outcomes Study 36-Item Short Form; SE, standard error; CI, confidence interval.

Figure 1b. Standardized Response Mean (SRM) of the SF-36 Mental Component Summary (MCS) Score

Review: SF-36
Comparison: 02 SRM
Outcome: 02 MCS

Study or subcategory	SRM (SE)	Weight (%)	SRM, random (95% CI)
Cohen 2006[37]	0.3345 (0.0900)	20.60	0.33 (0.16, 0.51)
Kosinski 2000[38]	0.1139 (0.1044)	15.31	0.11 (–0.09, 0.32)
Lipsky 2000a[40]	0.0726 (0.2390)	2.92	0.07 (–0.40, 0.54)
Lipsky 2000b[40]	0.0863 (0.2390)	2.92	0.09 (–0.38, 0.55)
Lipsky 2000c[40]	0.0899 (0.2387)	2.93	0.09 (–0.38, 0.56)
Lipsky 2000d[40]	0.0811 (0.2405)	2.88	0.08 (–0.39, 0.55)
Wells 2007a[43]	0.1297 (0.0830)	24.22	0.13 (–0.03, 0.29)
Wells 2007b[43]	0.3308 (0.1089)	14.07	0.33 (0.12, 0.54)
Zhao 2000a[45]	0.2270 (0.2162)	3.57	0.23 (–0.20, 0.65)
Zhao 2000b[45]	0.2973 (0.2160)	3.58	0.30 (–0.13, 0.72)
Zhao 2000c[45]	0.2212 (0.2182)	3.50	0.22 (–0.21, 0.65)
Zhao 2000d[45]	0.2477 (0.2182)	3.50	0.25 (–0.18, 0.68)
Total (95% CI)		100.0	0.21 (0.3, 0.29)



Test for heterogeneity: $Chi^2 = 6.27$, df = 11 ($P = .85$), $I^2 = 0\%$.
Test for overall effect: Z = 5.12 ($P < .00001$).

a, b, c, and d = treatment arms analyzed.
SF-36 indicates Medical Outcomes Study 36-Item Short Form; SE, standard error; CI, confidence interval.

Westhovens et al, 2006. "Improved Health-related Quality of Life for Rheumatoid Arthritis Patients Treated With Abatacept Who Have Inadequate Response to Anti-TNF Therapy in a Double-blind, Placebo-controlled, Multicentre Randomized Clinical Trial."[38] The overall objective of this clinical trial was to demonstrate the effect of abatacept on QOL for RA patients with inadequate response to anti-TNF therapy. QOL was assessed using the SF-36, HAQ, and fatigue visual analog scale. The active-treatment group QOL scores improved significantly on the HAQ and fatigue indices with improvement in 7 of 8 SF-36 scale components and SF-36 PCS and MCS scores. Improvement rate for QOL was faster for abatacept than for placebo, and the improvements from abatacept returned patients to normal levels of QOL on many domains.

Zhao et al, 2000. "Evaluation of Health-related Quality of Life of Rheumatoid Arthritis Patients Treated With Celecoxib."[45] Patients with diagnosed and active RA were enrolled and randomly assigned to 1 of 5 treatment groups: placebo, twice-daily celecoxib 100 mg, twice-daily celecoxib 200 mg, twice-daily celecoxib 400 mg, and twice-daily naproxen 500 mg. HAQ data were scored according to developers' specifications, with a low score representing a better functional status. The SF-36 data were also scored according to developers' specifications. Celecoxib was better than placebo and comparable with naproxen in improving functional status and overall HRQOL among RA patients.

The SRMs and 95% confidence intervals for the outcomes are provided in **Table 4** and the forest plots are displayed in **Figures 1a** and **1b**. Moderate effect size (ie, 0.50 ± 0.10) was found for PCS [0.42] and the HAQ [0.46] with a low effect size [ie, 0.20 ± 0.10] for the MCS [0.21]. The REs in relation to the TJC are also shown in Table 3. The overall REs were 0.96 for HAQ, 0.8 for PCS, and 0.2 for MCS, again showing better responsiveness for HAQ and PCS compared with the MCS, as expected.

Discussion

This article has summarized the evidence for the responsiveness to change of the SF-36 component scores across 7 placebo-controlled trials as assessed by SRM and effect size—2 key measures for meeting the OMERACT criteria for responsiveness to change. As expected, the PCS showed more responsiveness to change from nonsteroidal anti-inflammatory drugs and biologics than did the MCS, given that mobility and pain are more heavily weighted in the PCS. The PCS performed in a similar fashion to the HAQ, so that sample sizes adequate for the HAQ can be expected to have sufficient power for the PCS. However, much larger sample sizes would be needed to achieve adequate statistical power if the MCS is an important outcome.

Others have assessed responsiveness to change of the SF-36 components but have only studied 1 or 2 trials and, with the exception of Wells et al,[39] have used other indices. In 2 trials of misoprostol and diclofenac sodium versus placebo, Kosinski et al[41] found that changes in the SF-36 and HAQ scores were more strongly related to changes in the patient and physician global assessments and patient pain assessment than to changes in the joint swelling and tenderness counts.

Eichler et al,[13] in 2 placebo-controlled clinical trials that compared etoricoxib, naproxen, and placebo in 1684 patients groups, found that although the correlation with the joint scores was low, the association of clinical efficacy end points was nearly identical for the HAQ overall score and the SF-36 PCS.

Tuttleman et al[14] evaluated the SF-36 as a generic functional health status measure in 207 patients in the Minocycline in Rheumatoid Arthritis trial. The SF-36 had high internal consistency and reliability and high discriminant and convergent validity. Moderate correlations were observed for comparable items on the SF-36 and MHAQ regarding dressing, walking, and bending. Joint tenderness score correlations with items on the MHAQ and SF-36 scales were higher than for joint swelling scores. Physician and patient global assessments were most highly correlated with the SF-36 bodily pain item. Based on the data from this study, the authors have confirmed that the SF-36 is a valid instrument for patients with RA and that the SF-36 correlates with the MHAQ and the physician and patient global assessments.

These studies show that the SF-36 deserves serious consideration for inclusion in the core set of outcomes recommended for future trials. This will also expand the database on its performance. As a generic QOL measure, the SF-36 is better suited to capture the holistic health of the patient as reflected in the World Health Organization definition of health as being not only the avoidance of disease but the physical, emotional, and social well-being of the patient. Furthermore, use of the SF-36 permits comparisons of physical and mental aspects of QOL in the RA patient population, as well as comparisons of QOL parameters between patients with RA, other patient groups, and the general population. This contribution is unique and of added value when issues of QOL are important.

Acknowledgments

We would like to thank the following individuals for providing us with original data: Mark Kosinksi, Sarah Hewlett, Barbara Tilley, and Marilyn Tuttleman. We would also like to thank Mark Kosinski for allowing us access to his database of more than 10 000 studies on the SF-36.

REFERENCES

1. Preamble to the Constitution of the World Health Organization as adopted by the International Health Conference, New York, 19-22 June, 1946; signed on 22 July 1946 by the representatives of 61 States (Official records of the World Health Organization, no. 2, p. 100) and entered into force on 7 April 1948. New York, NY: 1948.

2. Murray CJL, Lopez AD. *The Global Burden of Disease: A Comprehensive Assessment of Mortality and Disability from Diseases, Injuries, and Risk Factors in 1990 and Projected to 2020.* Cambridge, MA: Harvard School of Public Health; 1996.

3. Spitzer WO. State of science 1986: quality of life and functional status as target variables for research. *J Chronic Dis.* 1987;40:465-471.

4. Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. *Health Technol Assess.* 1998;2:i-iv, 1-74.

5. Russak S, Croft J Jr, Furst D, et al. The use of rheumatoid arthritis health-related quality of life patient questionnaires in clinical practice: lessons learned. *Arthritis Rheum.* 2003;49:574-584.

6. Fries JF, Spitz PW, Young DY. The dimensions of health outcomes: the health assessment questionnaire, disability and pain scales. *J Rheumatol.* 1982;9:789-793.

7. McDowell I. *Measuring Health: A Guide to Rating Scales and Questionnaires.* 3rd ed. New York, NY; Oxford University Press; 2006.

8. Ware JE, Sherbourne CD. The MOS 36-Item Short-Form Health Survey (SF-36). I. Conceptual framework and item selection. *Med Care.* 1992;30:473-483.

9. What is OMERACT? www.omeract.org. Accessed October 31, 2007.

10. Boers M, Brooks P, Simon LS, Strand V, Tugwell P. OMERACT: an international initiative to improve outcome measurement in rheumatology. *Clin Exp Rheumatol.* 2005;23(5 suppl 39):S10-S13.

11. Brooks P, Boers M, Simon L, Strand V, Tugwell P. Outcome measures in rheumatoid arthritis-the OMERACT process. *Expert Rev Clin Immunol.* 2007;3:271-275.

12. Boers M, Brooks P, Strand CV, Tugwell P. The OMERACT filter for outcome measures in rheumatology. *J Rheumatol.* 1998;25:198-199.

13. Eichler HG, Mavros P, Geling O, Hunsche E, Kong S. Association between health-related quality of life and clinical efficacy endpoints in rheumatoid arthritis patients after four weeks treatment with anti-inflammatory agents. *Int J Clin Pharmacol Ther.* 2005; 43:209-216.

14. Tuttleman M, Pillemer SR, Tilley BC, et al. A cross sectional assessment of health status instruments in patients with rheumatoid arthritis participating in a clinical trial. Minocycline in Rheumatoid Arthritis Trial Group. *J Rheumatol.* 1997;24:1910-1915.

15. Bilberg A, Ahlmén M, Mannerkorpi K. Moderately intensive exercise in a temperate pool for patients with rheumatoid arthritis: a randomized controlled study. *Rheumatology (Oxford).* 2005;44: 502-508.

16. Cohen S, Cannon GW, Schiff M, et al. Two-year, blinded, randomized, controlled trial of treatment of active rheumatoid arthritis with leflunomide compared with methotrexate. Utilization of Leflunomide in the Treatment of Rheumatoid Arthritis Trial Investigator Group. *Arthritis Rheum.* 2001;44:1984-1992.

17. Genovese MC, Becker JC, Schiff M, et al. Abatacept for rheumatoid arthritis refractory to tumor necrosis factor alpha inhibition. *N Engl J Med.* 2005;353:1114-1123.

18. Harrison MJ, Kim CA, Silverberg M, Paget SA. Does age bias the aggressive treatment of elderly patients with rheumatoid arthritis? *J Rheumatol.* 2005;32:1243-1248.

19. Helliwell PS, O'Hara M, Holdsworth J, Hesselden A, King T, Evans P. A 12-month randomized controlled trial of patient education on radiographic changes and quality of life in early rheumatoid arthritis. *Rheumatology (Oxford).* 1999;38:303-308.

20. Hewlett S, Kirwan J, Pollock J, et al. Patient initiated outpatient follow up in rheumatoid arthritis: six year randomised controlled trial. *BMJ.* 2005;330:171.

21. Kaplan RM, Groessl EJ, Sengupta N, Sieber WJ, Ganiats TG. Comparison of measured utility scores and imputed scores from the SF-36 in patients with rheumatoid arthritis. *Med Care.* 2005; 43:79-87.

22. Keystone EC, Kavanaugh AF, Sharp JT, et al. Radiographic, clinical, and functional outcomes of treatment with adalimumab (a human anti-tumor necrosis factor monoclonal antibody) in patients with active rheumatoid arthritis receiving concomitant methotrexate therapy: a randomized, placebo-controlled, 52-week trial. *Arthritis Rheum.* 2004;50:1400-1411.

23. Kosinski M, Kuyawski S, Martin R. Health-related quality of life in early rheumatoid arthritis: impact of disease and treatment response. *Am J Manag Care.* 2002;8:231-240.

24. Martin LJ, Griffith SM. High disease activity scores predict the need for additional health services in patients over 60 with rheumatoid arthritis. *Musculoskeletal Care.* 2006;4:1-11.

25. Mathias SD, Colwell HH, Miller DP, Moreland LW, Buatti M, Wanke L. Health-related quality of life and functional status of patients with rheumatoid arthritis randomly assigned to receive etanercept or placebo. *Clin Ther.* 2000;22:128-139.

26. Scott DL. Leflunomide improves quality of life in rheumatoid arthritis. *Scand J Rheumatol Suppl.* 1999;112:23-29.

27. Scott DL, Garrood T. Quality of life measures: use and abuse. *Baillieres Best Pract Res Clin Rheumatol.* 2000;14:663-687.

28. Sköldstam L, Hagfors L, Johansson G. An experimental study of Mediterranean diet intervention for patients with rheumatoid arthritis. *Ann Rheum Dis.* 2003;62:208-214.

29. Strand V. Longer term benefits of treating rheumatoid arthritis: assessment of radiographic damage and physical function in clinical trials. *Clin Exp Rheumatol.* 2004;22(5 suppl 35): S57-S64.

30. Strand V, Cohen S, Crawford B, Smolen JS, Scott DL, for the Leflunomide Investigators Group. Patient-reported outcomes better discriminate active treatment from placebo in randomized controlled trials in rheumatoid arthritis. *Rheumatology (Oxford).* 2004;43:640-647.

31. Strand V, Scott DL, Emery P, et al. Physical function and health related quality of life: analysis of 2-year data from randomized, controlled studies of leflunomide, sulfasalazine, or methotrexate in patients with active rheumatoid arthritis. *J Rheumatol.* 2005;32:590-601.

32. Tijhuis GJ, Zwinderman AH, Hazes JM, Breedveld FC, Vlieland PM. Two-year follow-up of a randomized controlled trial of a clinical nurse specialist intervention, inpatient, and day patient team care in rheumatoid arthritis. *J Adv Nurs.* 2003;41:34-43.

33. Torrance GW, Tugwell P, Amorosi S, Chartash E, Sengupta N. Improvement in health utility among patients with rheumatoid arthritis treated with adalimumab (a human anti-TNF monoclonal antibody) plus methotrexate. *Rheumatology (Oxford).* 2004;43: 712-718.

34. Tugwell P, Wells G, Strand V, et al. Clinical improvement as reflected in measures of function and health-related quality of life following treatment with leflunomide compared with methotrexate in patients with rheumatoid arthritis: sensitivity and relative efficiency to detect a treatment effect in a twelve-month, placebo-

controlled trial. Leflunomide Rheumatoid Arthritis Investigators Group. *Arthritis Rheum.* 2000;43:506-514.

35. Weinblatt ME, Keystone EC, Furst DE, et al. Adalimumab, a fully human anti-tumor necrosis factor alpha monoclonal antibody, for the treatment of rheumatoid arthritis in patients taking concomitant methotrexate: the ARMADA trial. *Arthritis Rheum.* 2003;48:35-45.

36. Wolfe F, Michaud K. Towards an epidemiology of rheumatoid arthritis outcome with respect to treatment: randomized controlled trials overestimate treatment response and effectiveness. *Rheumatology (Oxford).* 2005;44(suppl 4):iv18-iv22.

37. Strand V, Tugwell P, Bombardier C, et al. Function and health-related quality of life: results from a randomized controlled trial of leflunomide versus methotrexate or placebo in patients with active rheumatoid arthritis. Leflunomide Rheumatoid Arthritis Investigators Group. *Arthritis Rheum.* 1999;42:1870-1878.

38. Westhovens R, Cole JC, Li T, et al. Improved health-related quality of life for rheumatoid arthritis patients treated with abatacept who have inadequate response to anti-TNF therapy in a double-blind, placebo-controlled, multicentre randomized clinical trial. *Rheumatology (Oxford).* 2006;45:1238-1246.

39. Wells GA, Li T, Maxwell L, Maclean R, Tugwell P. Responsiveness of patient-reported outcomes including fatigue, sleep quality, activity limitation, and quality of life following treatment with abatacept in patients with rheumatoid arthritis *Ann Rheum Dis.* 2007; [Epub ahead of print].

40. Cohen SB, Emery P, Greenwald MW, et al. Rituximab for rheumatoid arthritis refractory to anti-tumor necrosis factor therapy: results of a multicenter, randomized, double-blind, placebo-controlled, phase III trial evaluating primary efficacy and safety at twenty-four weeks. *Arthritis Rheum.* 2006;54:2793-2806.

41. Kosinski M, Zhao SZ, Dedhiya S, Osterhaus JT, Ware JE Jr. Determining minimally important changes in generic and disease-specific health-related quality of life questionnaires in clinical trials of rheumatoid arthritis. *Arthritis Rheum.* 2000;43:1478-1487.

42. Kremer JM, Genovese MC, Cannon GW, et al. Concomitant leflunomide therapy in patients with active rheumatoid arthritis despite stable doses of methotrexate. A randomized, double-blind, placebo-controlled trial. *Ann Intern Med.* 2002;137:726-733.

43. Lipsky PE, van der Heijde DM, St Clair EW, et al. Infliximab and methotrexate in the treatment of rheumatoid arthritis. Anti-Tumor Necrosis Factor Trial in Rheumatoid Arthritis with Concomitant Therapy Study Group. *N Engl J Med.* 2000;343:1594-1602.

44. Maini RN, Breedveld FC, Kalden JR, et al. Sustained improvement over two years in physical function, structural damage, and signs and symptoms among patients with rheumatoid arthritis treated with infliximab and methotrexate. *Arthritis Rheum.* 2004;50:1051-1065.

45. Zhao SZ, Fiechtner JI, Tindall EA, et al. Evaluation of health-related quality of life of rheumatoid arthritis patients treated with celecoxib. *Arthritis Care Res.* 2000;13:112-121.

46. Simon LS. Arthritis: new agents herald more effective symptom management. *Geriatrics.* 1999;54:37-42; quiz 44.