Treatment of Rheumatoid Arthritis Patients with Abatacept and Methotrexate Significantly Improved Health-Related Quality of Life

PAUL EMERY, MARK KOSINSKI, TRACY LI, MARIE MARTIN, G. RHYS WILLIAMS, JEAN-CLAUDE BECKER, BONNIE BLAISDELL, JOHN E. WARE Jr, CHARLES BIRBARA, and ANTHONY S. RUSSELL

ABSTRACT. ***Objective.*** This study examined the effect of abatacept, a costimulation modulator, on the health-related quality of life (HRQOL) of patients with rheumatoid arthritis (RA).

Methods. Three hundred thirty-nine patients with RA on a background of methotrexate (MTX), who participated in a multicenter, double-blind, placebo-controlled trial, were randomized to abatacept 2 mg/kg, abatacept 10 mg/kg, or placebo. HRQOL was assessed at pretreatment, and at 3, 6, and 12 months posttreatment using the SF-36 Health Survey (SF-36). Changes in SF-36 scores from baseline to 12 months were compared across treatment and placebo groups to examine HRQOL benefits of abatacept. A link between American College of Rheumatology improvement and changes in SF-36 scores was established to demonstrate the association between HRQOL outcomes and clinical response.

Results. After 12 months of treatment, patients randomized to abatacept 10 mg/kg showed significantly better HRQOL outcomes overall versus patients randomized to placebo (MANOVA $F = 4.71$, $p < 0.001$) or to abatacept 2 mg/kg (MANOVA $F = 1.97$, $p = 0.05$). Differences in SF-36 change scores between abatacept 10 mg/kg and placebo groups reached statistical significance on all 8 domain scales, the 2 summary measures, and the SF-36 utility index (SF-6D). Differences in SF-36 change scores between abatacept 10 mg/kg and abatacept 2 mg/kg reached statistical significance on 5 of the 8 domain scales, the physical summary measure, and the SF-6D. Improvement in HRQOL was highly related to clinical response.

Conclusion. Abatacept 10 mg/kg plus MTX demonstrated a stronger HRQOL response than placebo plus MTX. The abatacept 2 mg/kg arm showed a very weak and transient response. (First Release Mar 1, 2006; J Rheumatol 2006; 33:681–9)

Key Indexing Terms:
HEALTH-RELATED QUALITY OF LIFE | RHEUMATOID ARTHRITIS | SF-36
HEALTH OUTCOMES | HEALTH ASSESSMENT | ABATACEPT

A chronic and progressive disease such as rheumatoid arthritis (RA), characterized by joint pain, stiffness, and joint deformity as well as by varying degrees of physical impairment, fatigue, fever, and reactive depression[1], places a tremendous burden on the patient, their families, the healthcare system, and society at large. For patients, early RA (average disease duration < 18 mo) places a substantial burden on their physical functioning and emotional well-being that is comparable to diabetes or congestive heart failure[2]. As the disease progresses, patients experience increasing functional impairment, which may lead to work disability and lost wages. For the families of patients with RA, the progression of RA is likely to place a significant burden, particularly to the extent that the patient is physically impaired, in pain, emotionally distressed, and unable to work.

For the healthcare system that attends to patients with RA, the costs of office visits, medications, surgeries, hospitalizations, occupational therapy, and other social services amount to billions of dollars per year, with direct medical care costs in 2001 for an RA patient estimated at US $9519 per year[3]. For society at large, the functional impairment that characterizes later stages of RA often results in work disability and lost opportunity to gain from the RA patient's contribution to the workforce. Estimates place the amount of lost wages due to RA at US $2.5 billion per year[4].

From the Rheumatology and Rehabilitation Research Unit, University of Leeds, Leeds, UK; QualityMetric Incorporated, Lincoln, Rhode Island, USA; Bristol-Myers Squibb, Princeton, New Jersey, USA; Health Assessment Lab, Boston, Massachusetts, USA; Tufts University School of Medicine, Boston, MA, USA; University of Massachusetts, Worcester, MA, USA; and Department of Medicine, University of Alberta, Edmonton, Alberta, Canada.

Supported by a grant from Bristol-Myers Squibb.

P. Emery, MA, MD, FRCP, Professor, Rheumatology and Rehabilitation Research Unit, University of Leeds; M. Kosinski, MA, Senior Scientist, QualityMetric Inc.; T. Li, PhD, Associate Director, Immunology, Bristol-Myers Squibb; M. Martin, PhD, Director, Consulting Services, QualityMetric Inc.; G.R. Williams, ScD, Director, Outcomes Research; J-C. Becker, MD, Clinical Director, Bristol-Myers Squibb; B. Blaisdell, MA, Outcomes Research, QualityMetric Inc.; J.E. Ware Jr, PhD, CEO, Chief Science Officer, Chairman of the Board, QualityMetric Inc., Health Assessment Lab, Tufts University School of Medicine; C. Birbara, MD, Rheumatologist, University of Massachusetts; A.S. Russell, MD, Professor of Rheumatology, Department of Medicine, University of Alberta.

Address reprint requests to Prof. P. Emery, Department of Rheumatology and Rehabilitation, University of Leeds, 36 Clarendon Road, Leeds LS2 9N2, UK. E-mail: p.emery@leeds.ac.uk

Accepted for publication November 25, 2005.

Pharmacologic interventions that improve physical function so that patients are able to work longer and perform daily activities have the potential to improve patients' quality of life as well as reduce the burden that RA places on society as a whole.

Abatacept, a novel selective T cell costimulation modulator, has been shown to be safe, well tolerated, and able to produce a significant dose-dependent reduction in disease activity for RA patients who experienced inadequate responses to disease modifying antirheumatic drugs (DMARD)[5,6]. Recent results from a randomized, controlled study showed that abatacept significantly improved the signs and symptoms of disease in patients who had active RA despite ongoing methotrexate (MTX) treatment[5]. In this study, we examined the effect of abatacept therapy (10 mg/kg) in combination with MTX on a broad range of health-related quality of life (HRQOL) domains in patients who had inadequate response to MTX treatment.

MATERIALS AND METHODS

Study population. Three hundred thirty-nine patients with RA participated in a multicenter, multinational, double-blind, randomized, placebo-controlled trial comparing the efficacy of MTX plus placebo (n = 119), abatacept 2 mg/kg (n = 105), or abatacept 10 mg/kg (n = 115). Abatacept was administered by intravenous infusions at baseline, every 2 weeks for the first month, and monthly thereafter. To participate, patients were required to meet several criteria: (1) the American Rheumatism Association criteria for RA while meeting functional class I, II, or III according to the revised criteria of the American College of Rheumatology (ACR)[7,8]; (2) have > 10 swollen, > 12 tender joints, and C-reactive protein level > 1 mg/dl signifying active disease; (3) have been treated with MTX for at least 6 months and on a stable dose for 28 days prior to enrollment; and (4) be washed-out of all DMARD other than MTX for at least 28 days before treatment. Provided that the prescribed dose remained stable for the first 6 months of the study, participants were permitted to continue on low-dose corticosteroids (≤ 10 mg/day) and nonsteroidal antiinflammatory drugs. This study was carried out in accordance with the ethical principles of the Declaration of Helsinki.

General health status measures. The Medical Outcomes Study Short Form-36 Health Survey (SF-36), a well-validated measure of general health status[9-11], was self-administered at baseline (pretreatment) and 3, 6, and 12 months posttreatment to measure HRQOL. The SF-36 measures 8 health dimensions [physical functioning (PF), role limitations due to physical health (RP), bodily pain (BP), general health perceptions (GH), vitality (VT), social functioning (SF), role limitations due to emotional health (RE), mental health (MH)], which are aggregated to produce physical (PCS) and mental (MCS) summary measures. All SF-36 scales and both summary measures were scored using norm-based methods that standardize the scores to a mean of 50 and a standard deviation of 10 in the general US population, with higher scores indicative of better health[12,13]. A health utility index (SF-6D) was also derived from 11 items of the SF-36[14]. The SF-6D is a preference-based measure of health that places the observable states of health and functioning on a preference continuum with a value of 0 for death through to 1 for completely well or optimal health. These values are known as health state utilities, which are primarily used to adjust life-years saved by quality for use in economic evaluations and decision models.

Effect of treatment on HRQOL. HRQOL outcomes were evaluated in 2 ways. First, changes in SF-36 scale and summary measure scores and the SF-6D from baseline to 12 months were evaluated and compared between placebo and treatment groups. To account for multiple comparisons, multivariate analysis of variance (MANOVA) was conducted to test for overall differences in change scores across all 8 SF-36 scales between placebo and abatacept groups. MANOVA analyses were followed with independent pairwise t tests of statistically significant differences in mean change scores on each SF-36 scale, summary measure, and the SF-6D between placebo and abatacept groups. For these analyses, the last observation was carried forward for those patients who dropped out early from the study.

Since presenting mean changes in scores can mask the underlying variability in HRQOL outcomes, the second way that HRQOL outcomes were evaluated was to determine the percentage of patients in each group whose 12 month score on each SF-36 scale was "better," "the same," or "worse" than the baseline score. Given that standards for the minimal clinically important difference for the SF-36 have not been well established, the standard error of measurement (SEM) was calculated for each SF-36 scale and summary measure, which provided a boundary of measurement error that could be used to categorize changes in individual patient scores as better, the same, or worse. The SEM is a theoretically fixed characteristic of a measurement instrument incorporating a sample's variability (σ_x) and reliability (r_{xx}), simply defined as $\sigma_x(1 - r_{xx})$ [15], and has been used to define the minimal clinically important difference (MCID) of several HRQOL measures[16,17]. We used variance and reliability estimates from a representative sample of the general US population for the SF-36[12], since the lack of variance in scores observed at baseline in the trial population could effectively reduce the size of the estimated SEM[18]. Two SEM, which corresponds to the 95% confidence interval around an individual patient score, was used to categorize patients as better, the same, or worse. In other words, there was only a one in 20 chance that we could incorrectly classify a patient as better, the same, or worse given measurement error. For the SF-6D, the classification of patients into better, same, or worse groups was based on the MCID of 0.041 points established in a previous study[19]. Chi-square tests were conducted to determine if significant differences existed in the proportion of RA patients whose 12-month HRQOL was "better," "the same," or "worse" than baseline between placebo and treatment groups.

Content-based interpretation of results. In an effort to provide meaningful interpretation of the differences in HRQOL outcomes observed between placebo and treatment groups we examined the responses to the following questionnaire items from each of the 9 SF-36 scales: (1) walking one block (PF); (2) cut down time spent at work because of physical health (RP); (3) pain interference (BP); (4) general health rating (GH); (5) feeling tired or worn out (VT); (6) health interferes with social activities; (7) doing work less carefully because of emotional problems (RE); and (8) feeling downhearted and blue. These items were selected because they provide the most salient interpretation of score changes on each of their respective scales. For this analysis we derived a dichotomous variable for each of the 8 SF-36 items. For example, the PF item on walking one block asks the patient to indicate the extent of limitations in walking one block. The response options for this item include "yes, limited a lot," "yes, limited a little," and "no, not limited at all." We examined the percentage of patients who indicated "any" limitations (yes, limited a lot or yes, limited a little) in walking a block at baseline and at 12 months as a way of understanding the meaning of the difference in change scores on the SF-36 physical functioning scale between the placebo and treatment groups. A similar strategy was used for evaluating the other 7 items selected for this analysis.

Linking HRQOL outcomes to clinical response. Another measure taken to improve the interpretation of HRQOL outcomes observed in this study was to link changes in SF-36 scale and summary measure scores and the SF-6D from baseline to 12 months to ACR improvement. All trial participants were placed into 4 mutually exclusive groups of ACR improvement: (1) ACR improvement < 20%; (2) ACR improvement from 20% to 49%; (3) ACR improvement from 50% to 69%; and (4) ACR improvement of 70% or greater. Mean changes in SF-36 scale and summary measure scores and SF-6D were compared across these 4 groups using ANOVA methods.

RESULTS

Patient characteristics. Baseline demographics, disease characteristics, and HRQOL summary scores of patients by treat-

ment group and sample disposition are presented in Table 1. Demographically, the 2 groups were very similar. The disease characteristics of both groups of patients were mixed, where in some cases the patients who discontinued showed more active disease and in other cases the patients who continued showed more active disease. Lastly, baseline HRQOL was similar in terms of physical health status. However, differences were observed in 2 of the 3 groups (placebo and abatacept 10 mg), where patients who discontinued showed worse mental health status.

Baseline demographics, disease characteristics, and HRQOL scores of all 339 patients enrolled in the study are presented in Table 2. Patients had a mean age of 55 years (range 17 to 83), 68% were female, 87% were Caucasian, 82% were rheumatoid factor-positive, and the mean duration of RA was 9–10 years. Baseline mean number of tender and swollen joints was similar across the treatment groups. Despite longterm treatment with MTX, patients had a high degree of disease activity at baseline (mean tender joint count ranged from 28.2 to 30.8 and mean swollen joint count ranged from 20.2 to 21.8). As expected, given the high degree of disease activity observed at baseline, HRQOL scores at baseline

Table 1. Patient characteristics by sample disposition.

Demographic Variable*	Placebo + MTX		Abatacept 2 mg/kg + MTX		Abatacept 10 mg/kg + MTX	
	Continued	Discontinued	Continued	Discontinued	Continued	Discontinued
Sample size, n	71	48	74	31	90	25
Age, yrs, mean (SD)	54.0 (12)	55.6 (12)	54.4 (11)	54.5 (11)	56.1 (13)	54.8 (11)
Female, %	69.0	63.0	62.0	65.0	73.0	80.0
Caucasian, %	87.0	88.0	84.0	94.0	88.0	84.0
MTX dose, mean (SD)	15.8 (4)	16.0 (4)	15.5 (5)	16.3 (4)	14.9 (5)	15.5 (4)
RA duration, yrs, mean (SD)	7.9 (8)	10.5 (8)	8.8 (8)	11.6 (7)	10.1 (10)	8.6 (8)
Rheumatoid factor + (%)	77.0	73.0	84.0	90.0	87.0	84.0
Tender joint count, mean	29.3	29.1	26.6	32.0	30.6	31.5
Swollen joint count, mean	22.1	21.4	19.1	23.0	20.8	23.0
Mean baseline HRQOL (SD)						
SF-36 physical summary	32.7 (8)	31.5 (7)	31.0 (9)	29.8 (7)	30.9 (8)	30.1 (9)
SF-36 mental summary	44.2 (12)	39.3 (12)	42.5 (13)	46.5 (12)	46.2 (12)	43.4 (13)

RA: rheumatoid arthritis; MTX: methotrexate. Continued: patients who completed the trial. Discontinued: patients who discontinued prior to 12 mo and whose last observation was carried forward for analyses. * All demographics at baseline are nonsignificant.

Table 2. Patient characteristics of analytic sample[†].

Demographic Variable*	Placebo + MTX	Abatacept 2 mg/kg + MTX	Abatacept 10 mg/kg + MTX
Sample size, n	119	105	115
Age, yrs, mean (range)	54.7 (23–80)	54.4 (23–80)	55.8 (17–83)
Female, %	66.4	62.9	74.8
Caucasian, %	87.4	86.7	86.9
MTX dose, mean (SD)	15.8 (4.1)	15.8 (4.8)	15.0 (4.4)
RA duration, yrs, mean (SD)	8.9 (8.3)	9.7 (8.1)	9.7 (9.8)
Rheumatoid factor +, %	75.6	85.7	86.1
Tender joint count, mean	29.2	28.2	30.8
Swollen joint count, mean	21.8	20.2	21.3
Mean baseline HRQOL score (SD)			
SF-36 physical functioning	30.4 (10.0)	29.7 (9.7)	29.7 (9.7)
SF-36 role physical	33.5 (8.8)	32.3 (8.0)	33.5 (9.4)
SF-36 bodily pain	35.2 (7.9)	34.3 (7.7)	35.4 (8.4)
SF-36 general health	36.9 (8.8)	36.4 (9.3)	37.0 (9.2)
SF-36 vitality	40.2 (9.8)	40.2 (8.8)	39.6 (8.8)
SF-36 social functioning	36.7 (11.0)	35.8 (10.9)	38.5 (11.6)
SF-36 role emotional	36.2 (13.9)	38.4 (14.6)	40.0 (14.2)
SF-36 mental health	41.6 (12.5)	42.3 (12.1)	44.1 (12.1)
SF-36 physical summary	32.2 (7.5)	30.7 (8.5)	30.7 (8.4)
SF-36 mental summary	42.2 (12.6)	43.6 (12.8)	45.6 (12.6)
SF-36 utility index (SF-6D)	0.55 (0.11)	0.56 (0.10)	0.57 (0.11)

RA: rheumatoid arthritis; MTX: methotrexate. * All demographics at baseline are nonsignificant. [†] Includes patients who discontinued from the study prior to 12 mo and were carried forward from their last observation for analyses.

were well below general population norms[12], particularly those HRQOL scales measuring physical health status (SF-36 physical functioning, role physical, and bodily pain scales). No significant differences in baseline patient characteristics or HRQOL scores were detected across treatment groups.

Effect of treatment on HRQOL (Table 3). Overall, the changes in SF-36 scales and summary measure scores differed between abatacept 10 mg/kg and placebo group (MANOVA F = 4.7, $p < 0.0001$). Mean improvements from baseline ranged from 5.3 to 9.3 points across the 8 SF-36 scales, and were 8.0 and 5.7 points for the physical and mental summary measures, respectively. Results of independent t tests showed that the differences in mean change scores were statistically significant between abatacept 10 mg/kg and placebo groups on all 8 scales and both summary measures, with outcomes favoring the abatacept 10 mg/kg group. Differences in mean score changes on the SF-36 scale and summary measures between abatacept 10 mg/kg and placebo groups ranged from 2.5 to 5.8 points, with the largest differences observed on the SF-36 bodily pain (5.8 points), vitality (5.8 points), physical summary (5.4 points), and physical functioning (5.1 points) scales. A statistically significant difference in mean score change on the SF-6D (0.05 points) was also observed between abatacept 10 mg/kg and placebo groups, favoring the abatacept 10 mg/kg group.

Significant differences in HRQOL outcomes were also observed between the abatacept 10 mg/kg and 2 mg/kg groups (MANOVA F = 1.97, $p < 0.05$). Results of independent t tests conducted with each scale showed that mean changes were higher (better outcome) among patients in the abatacept 10 mg/kg group on 5 of the 8 SF-36 domains scales (PF, RP, BP, VT, and SF), the physical (PCS) summary measure, and the SF-6D compared to patients in the abatacept 2 mg/kg group. Differences in mean score changes between the 2 abatacept groups on these SF-36 scales ranged from 2.5 to 4.4 points, with the largest difference observed on the vitality scale (4.4 points). The difference in mean SF-6D score change between the 2 groups was 0.05 points.

Overall, differences between the abatacept 2 mg/kg and placebo groups in HRQOL outcomes did not differ significantly (MANOVA F = 1.42, $p = 0.19$), although results of independent t tests conducted for each SF-36 scale and summary measure showed a significant difference in mean score change between the placebo and abatacept 2 mg/kg group on the physical functioning (2.6 points) and bodily pain (3.0 points) scales and the physical summary measure (2.6 points), all favoring the abatacept 2 mg/kg group.

Analysis of change scores categorized as better, the same, or worse confirmed the results of the average patient in each treatment group, specifically that a greater proportion of patients in each group had better HRQOL outcomes than worse HRQOL outcomes (Table 4). Across all SF-36 scales and summary measures and the SF-6D a greater proportion of patients in the abatacept 10 mg/kg group improved more than would be expected due to measurement error versus patients in either the placebo group or abatacept 2 mg/kg group.

Table 3. Mean changes in HRQOL scores from baseline to 12 months.

	Mean Changes in Score from Baseline to 12 mo						Differences in Mean Change Scores†		
	Placebo + MTX		Abatacept 2 mg/kg + MTX		Abatacept 10 mg/kg + MTX		Placebo vs Abatacept 2 mg/kg	Abatacept 2 mg/kg vs Abatacept 10 mg/kg	Placebo vs Abatacept 10 mg/kg
	Mean	SE	Mean	SE	Mean	SE			
SF-36 scale									
PF	2.1	0.8	4.7	0.9	7.2	0.8	2.6[a]	2.5[a]	5.1[d]
RP	4.1	1.0	5.3	1.1	8.2	1.0	1.2	2.9[a]	4.1[b]
BP	3.5	0.8	6.5	0.9	9.3	0.8	3.0[a]	2.8[a]	5.8[d]
GH	2.3	0.7	4.1	0.7	5.8	0.7	1.8	1.7	2.5[c]
VT	2.1	0.8	3.5	0.8	7.9	0.8	1.4	4.4[c]	5.8[d]
SF	3.6	0.9	4.6	1.0	7.6	0.9	1.0	3.0[a]	4.0[b]
RE	3.8	1.0	5.0	1.4	6.8	0.9	1.8	1.8	3.0[c]
MH	2.6	0.8	3.5	0.9	5.3	0.8	0.9	1.8	2.7[a]
SF-36 summary									
PCS	2.6	0.7	5.2	0.8	8.0	0.8	2.6[a]	2.8[a]	5.4[d]
MCS	2.8	0.9	3.5	1.0	5.7	0.9	0.7	1.9	2.9[a]
SF-36 utility index									
SF-6D	0.06	0.01	0.06	0.01	0.11	0.01	0.00	0.05[c]	0.05[c]

HRQOL: health-related quality of life; SF-36: Medical Outcomes Study Short-Form 36 Health Survey. PF: physical functioning; RP: role physical; BP: bodily pain; GH: general health; VT: vitality; SF: social functioning; RE: role emotional; MH: mental health; PCS: physical component summary; MCS: mental component summary; SF-6D: SF-36 Utility Index. [a] p , 0.05; [b] $p < 0.01$; [c] $p < 0.001$; [d] $p < 0.0001$. Multiple comparison tests (MANOVA): Abatacept 10 mg/kg vs placebo: MANOVA F = 4.71, $p < 0.001$. Abatacept 2 mg/kg vs placebo: MANOVA F = 1.42, $p = 0.19$. Abatacept 10 mg/kg vs abatacept 2 mg/kg: MANOVA F = 1.97, $p = 0.05$.

Table 4. Categories of change in HRQOL scores from baseline to 12 months by treatment group.

	1 Placebo + MTX			2 Abatacept 2 mg/kg + MTX			3 Abatacept 10 mg/kg + MTX			Chi-square Test		
	% Worse	% Same	% Better	% Worse	% Same	% Better	% Worse	% Same	% Better	1 vs 2	1 vs 3	2 vs 3
SF-36 scale												
PF	8.4	64.7	26.9	3.8	61.0	35.2	7.8	45.2	47.0	3.3	10.5[b]	5.9[a]
RP	13.5	49.5	37.0	6.7	51.0	42.3	9.6	35.6	54.8	2.8	7.5[a]	6.0[a]
BP	10.1	56.8	33.1	2.9	52.4	44.7	2.4	37.1	60.5	5.9[a]	12.9[b]	5.9[a]
GH	4.2	75.6	20.2	1.0	77.1	21.9	1.7	55.6	42.7	2.3	14.2[c]	11.3[b]
VT	11.8	61.3	26.9	12.4	52.4	30.8	2.6	47.0	50.4	2.1	17.4[d]	10.5[b]
SF	5.9	66.4	27.7	11.4	49.5	39.1	6.1	49.6	44.7	6.9[a]	7.4[a]	2.2
RE	11.8	58.0	30.2	13.0	54.0	33.3	8.7	54.8	36.5	1.6	1.3	1.8
MH	7.6	69.7	22.7	7.6	63.8	28.6	2.6	59.9	37.5	1.1	6.9[a]	2.9
SF-36 summary												
PCS	10.1	58.0	31.9	6.7	49.0	44.2	6.9	32.7	60.4	3.8	15.4[d]	5.9[a]
MCS	10.9	57.2	31.9	9.6	57.7	32.7	4.8	47.1	48.1	0.1	6.9[a]	1.0
SF-36 utility index												
SF-6D	9.3	42.4	48.3	9.9	43.6	46.5	7.7	26.7	65.6	0.1	5.9[c]	5.9[c]

PF: physical functioning; RP: role physical; BP: bodily pain; GH: general health; VT: vitality; SF: social functioning; RE: role emotional; MH: mental health; PCS: physical component summary; MCS: mental component summary; SF-6D: SF-36 Utility Index. [a] p , 0.05; [b] p < 0.01; [c] p < 0.001; [d] p < 0.0001.

Differences in the proportion of patients who got better, stayed the same, or got worse reached statistical significance on 10 of 11 comparisons between the abatacept 10 mg/kg and placebo groups, 2 of 11 comparisons between the abatacept 2 mg/kg and placebo groups, and 7 of 11 comparisons between the abatacept 10 mg/kg and abatacept 2 mg/kg groups. The categorical analyses call attention to the substantial variation of outcomes observed in all 3 groups of patients. For example, a sizeable proportion of patients in each group showed changes in scores that were either within measurement error (same) or indicative of worse HRQOL outcomes.

Content-based interpretation of results. The content of selected SF-36 questionnaire items was examined as a basis for interpreting the differences in the HRQOL outcomes observed between treatment groups. In Table 5, results obtained at baseline for each item are contrasted to results obtained at 12 months and can be used to help interpret the changes in scores on each domain scale in Table 3. For example, patients in the abatacept 10 mg/kg group improved 7.2 points on the SF-36 physical functioning scale. Underlying that improvement was a substantial decrease in the percentage of patients reporting any limitations in walking one block from 72% at baseline to 42% at 12 months posttreatment, more than a 40% reduction in limitations in walking a block. Comparatively, there was virtually no change in the percentage of placebo patients reporting limitations in walking a block from baseline (61%) to 12 months (62%). Similarly, as Table 5 shows, underlying the significant improvement in SF-36 scale scores for the abatacept 10 mg/kg group was a substantial decrease in the percentage of patients who reported: (1) having to cut down on time at work because of health (70% to 32%); (2) pain interfered with work quite a bit or extremely (51% to 11%); (3) fair or poor health in general (64% to 26%); (4) felt tired or worn out all or most of the time (38% to 14%); (5) health interfered with social activities quite a lot or extremely (32% to 9%); (6) working less carefully than usual (42% to 23%); and (7) feeling downhearted and blue all or most of the time (18% to 4%).

Table 5. Percentage of patients reporting limitations in functional status and well-being: content-based interpretations of changes in SF-36 scale scores.

SF-36 Scale	Selected Item Content as Dichotomized	Placebo BL	Placebo 12 mo	Abatacept 2 mg/kg + MTX BL	Abatacept 2 mg/kg + MTX 12 mo	Abatacept 10 mg/kg + MTX BL	Abatacept 10 mg/kg + MTX 12 mo
PF	Any limitations in walking one block	61	62	69	49	72	42
RP	Cut down amount of time spent on work/activities	66	46	74	45	70	32
BP	Pain interfered with work quite a bit or extremely	56	21	61	21	51	11
GH	Fair or poor rating of health in general	64	33	62	39	64	26
VT	Felt tired all or most of the time	39	31	40	20	38	14
SF	Health interferes with social activities quite a lot or extremely	35	16	39	14	32	9
RE	Did not do work/activities as carefully as usual	57	36	51	32	42	23
MH	Felt downhearted and blue all or most of the time	10	7	22	11	18	4

BL: baseline percentages. 12 mo: 12-month percentages.

Linking HRQOL outcomes to clinical response. Table 6 presents mean changes in SF-36 scale and summary measure scores and the SF-6D across 4 categories of ACR improvement. As shown, the magnitude of the mean score improvement on each SF-36 scale, summary measure, and the SF-6D increased incrementally with increasing levels of ACR improvement, and overall the differences in mean score improvement across ACR improvement were statistically significant. The largest changes in SF-36 scale and summary measure scores and the SF-6D were observed among the group of patients who reached or exceeded 70% ACR improvement. With few exceptions these patients improved by more than 10 points on average (range 8.6 to 15.4 points) on each SF-36 scale and summary measure and by 0.17 points on the SF-6D. The next largest change on each SF-36 scale and summary measure and the SF-6D was observed among patients whose ACR improvement ranged from 50% to 69%. For these patients, score improvement ranged from 5.4 to 10.9 points across SF-36 scales and summary measures and was 0.13 points on the SF-6D. Changes in SF-36 scale and summary measure scores ranged from 2.8 to 7.6 points and the SF-6D improved by 0.09 points among patients whose ACR improvement ranged from 20% to 49%. The smallest changes in scores were observed among patients who did not reach 20% ACR improvement. Changes in scores ranged from 1.4 to 3.1 points across SF-36 scales and summary measures, and was 0.03 on the SF-6D among these patients.

DISCUSSION

We sought to evaluate the effect of abatacept, a costimulator, on health-related quality of life in patients with rheumatoid arthritis. Patients enrolled into this study were in a highly active disease state at baseline (pretreatment), which was cause for significant HRQOL burden at the outset of this study. HRQOL scores, as measured by the SF-36, at baseline were well below general US population norms and rivaled norms of other serious medical conditions such as congestive heart failure and advanced type I diabetes[12]. As expected with RA, the most pronounced HRQOL burden was detected with SF-36 scales measuring physical health status (physical functioning, role physical, and bodily pain scales). On average, scores on these physical health scales were 1.5 to 2.0 standard deviation units below general population norms, which are large effect size differences, and well below the average scores of persons older than 85 years of age[12]. Interestingly, baseline SF-36 scores of patients in this study also showed substantial emotional burden associated with a highly active disease state. The average baseline scores on the SF-36 mental health scale and mental summary measure were at the cut-point score (< 42) for a first-stage screen for depression[12]. Lastly, at baseline, average SF-6D scores ranged from 0.55 to 0.57 across the 3 treatment groups, indicating that patients enjoyed roughly one-half of "optimal" health.

Patients treated with abatacept 10 mg/kg showed statistically significant improvement from baseline to 12 months across all 8 SF-36 domain scales, both physical and mental summary measures, and the SF-36 utility index in the moderate to large effect size range[20]. In results not reported in this article, significant improvement in SF-36 scores observed for the abatacept 10 mg/kg group was also evident at 3 months, with the magnitude of change in SF-36 scores close to the amount of change observed at 6 and 12 months. The implication of the 3 and 6 month findings is that such rapid response perceived by the patient can have a positive impact on treatment compliance, since patients perceive the benefits of treatment early on.

Table 6. Mean changes in SF-36 scale and summary measure scores across categories of ACR improvement.

	Categories of ACR Improvement								
	70%, n = 46		50–69%, n = 49		20–49%, n = 63		< 20%, n = 181		
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	F
SF-36 scale									
PF	12.3	9.1	7.7	8.7	5.8	9.5	1.4	7.4	25.4*
RP	13.5	13.3	10.5	10.9	5.6	10.6	2.7	9.0	17.6*
BP	15.4	9.6	10.9	8.6	7.6	8.8	2.5	7.1	39.5*
GH	10.7	9.1	6.5	7.1	4.1	6.4	1.7	6.2	23.8*
VT	10.7	9.7	7.2	7.3	4.1	7.6	2.3	8.5	14.4*
SF	10.3	11.8	10.6	9.8	6.4	9.3	2.2	9.1	15.6*
RE	12.1	15.2	8.2	12.0	3.5	13.4	3.1	14.6	6.1*
MH	8.6	9.9	5.4	8.3	3.3	9.1	2.2	9.0	6.6*
SF-36 summary									
PCS	13.5	9.0	9.4	8.3	6.3	8.8	1.7	6.4	37.2*
MCS	8.7	10.9	6.5	8.8	2.8	10.6	2.6	10.2	5.6*
SF-36 utility index									
SF-6D	0.17	0.13	0.13	0.11	0.09	0.09	0.03	0.09	28.5*

* $p < 0.001$.

The score changes from baseline to 12 months observed on the SF-36 for abatacept 10 mg/kg patients were equivalent to or larger than those observed in previous studies involving disease modifying treatments[6,21-31]. Using variance estimates from the general population, mean changes in all SF-36 scale and summary measure scales for the abatacept 10 mg/kg group exceeded one-half a standard deviation, which could be considered clinically important[32]. Further, the HRQOL benefits of abatacept 10 mg/kg as measured by the SF-36 have been linked to a variety of clinical, social, and economic benefits[12,13]. Relative to placebo, moderate effect size differences in outcomes were observed on the SF-36 physical functioning, bodily pain, and vitality scales and the physical summary measure for the abatacept 10 mg/kg group. Differences in outcomes between placebo and abatacept 10 mg/kg on these scales exceeded 5 points, which was more than one-half a standard deviation difference in outcomes. The differences in outcomes between abatacept 10 mg/kg and placebo on the remaining SF-36 scales and summary measures were in the small effect size range.

A feature of this study not found in most HRQOL treatment studies was the analysis of the underlying variability of HRQOL outcomes. Most HRQOL treatment studies present results in terms of mean changes in HRQOL scores, which can mask the differences in HRQOL outcomes observed across individual patients. Recognizing that not all patients would experience the average HRQOL outcome, each patient was categorized as better, the same, or worse depending upon the magnitude and direction of the change in score on each HRQOL scale from baseline to 12 months. The magnitude of change necessary to be categorized as better or worse was based on 2 standard errors of measurement (SEM), or the 95% confidence interval around an individual patient score. For all HRQOL scales, 2 SEM exceeded one-half a standard deviation, which is considered clinically important[25]. As the results of this analysis showed, a greater proportion of patients in each group showed meaningful improvement across all SF-36 scales and summaries and the SF-6D compared to meaningful decline. The ratio of patients who improved versus declined was largest in the abatacept 10 mg/kg group. With the exception of the SF-36 role emotional scale, the ratio of patients who improved versus declined was at least 6 to 1 in the abatacept 10 mg/kg group and as high as 25 to 1 on the SF-36 bodily pain and general health scales.

Our study featured the SF-36 utility index, the SF-6D. The primary use of utility measures like the SF-6D is to adjust life-years saved by quality for use in economic evaluations and decision models. Preference-based health state scores do not have natural units. Since health is a function of both length and quality of life, the quality-adjusted life-year (QALY) has been developed to combine the values of these 2 attributes of health into a single number. In this study we observed clinically meaningful improvement on the SF-6D in all 3 treatment groups as defined by the developer of the SF-6D and colleagues[33]. The improvement in SF-6D scores among patients in the abatacept 10 mg/kg group was about twice that of the other 2 groups, or 0.05 points, which can be interpreted as a difference of 5 more patients in perfect health for one year compared to placebo or abatacept 2 mg/kg groups.

It is noteworthy that a dose-related response on HRQOL was observed in this study. The abatacept 10 mg/kg group induced substantially greater improvements from baseline across all HRQOL measures, compared to the abatacept 2 mg/kg group or placebo. This finding, combined with prior reports on improvements on signs and symptoms at 6 and 12 months of the study[5,6], emphasized that abatacept 10 mg/kg is a viable treatment option for patients with inadequate response to MTX. The improvement on HRQOL from abatacept 10 mg/kg was both statistically significant and clinically meaningful. Results from content-based analysis on specific SF-36 questionnaire items further demonstrated the real-life meaning of the improvements to the patients on their ability to walk, participate in work or social activities, and maintain a healthy mental status.

This study provided examples of how HRQOL measures can be used to understand the burden of disease and the potential benefits of treatment. One interpretation strategy used to translate the HRQOL scores into more salient terms was the analysis of responses to individual questions of the SF-36. This strategy has been used to interpret disease burden and treatment outcomes for the SF-36 in previous studies[34,35]. In our study, patients scored well below general population norms on each SF-36 scale, as expected. However, the interpretation of that burden became more salient by analyzing the content of specific items from each scale. For example, the analysis of selected SF-36 items demonstrated that more than 60% of the patients in each group reported limitations in walking one block or having to cut down time spent at work because of physical health problems. Similarly, the analysis of the content of specific items improved our interpretation of the differences in outcomes between treatment groups. As was observed, there was a 40% reduction in the percentage of patients who reported limitations in walking a block (72% to 42%) and a 50% reduction in the percentage of patients who reported having to cut down the time spent at work because of physical health (70% to 32%) in the abatacept 10 mg/kg group. By comparison, there was no reduction in the percentage of placebo patients who reported limitations in walking a block, and a 33% reduction in placebo patients who reported having to cut down time spent at work because of physical health.

Results of the analysis of changes in HRQOL scores by clinical response as measured by the ACR improvement criteria showed that patients who experienced the greatest clinical response also showed the greatest score improvement on all SF-36 scales, summary measures, and the SF-6D. The magnitude of mean score improvement on each scale increased incrementally with increasing levels of ACR improvement.

However, the results showed that not all HRQOL concepts were equally responsive to ACR improvement. As expected, changes in scores on SF-36 scales measuring physical health status (physical summary, bodily pain, physical functioning, general health, and role physical scales) were more strongly related to differences in ACR improvement than score changes in SF-36 scales measuring mental health status (mental summary, mental health, and role emotional scales), as indicated by the magnitude of the F statistics. A strong relationship was also evident between ACR improvement and the SF-6D. The results of these analyses can be useful in determining the clinical meaningfulness of HRQOL outcomes. In particular, differences in mean score changes between the group of patients showing little or no improvement in disease activity (ACR < 20%) and the group of patients showing a minimum to moderate change in disease activity (ACR 20%–49%) could be considered minimal clinically important changes in outcomes on each SF-36 scale. In this regard the HRQOL outcomes observed for the abatacept 10 mg/kg group would be considered clinically meaningful.

In conclusion, the disease burden that active RA places on patients' physical and emotional HRQOL is substantial. Treatment of RA by abatacept plus MTX improves HRQOL across a wide range of measured domains. Patient-reported HRQOL improvements were greatest for those who received abatacept 10 mg/kg. Future investigation should examine the extent to which the HRQOL improvement at 12 months among RA patients receiving abatacept therapy is sustained through time.

REFERENCES

1. Karlson EW, Katz JN, Liang MH. Chronic rheumatic disorders. In: Spilker B, editor. Quality of life and pharmacoeconomics in clinical trials. Philadelphia: Lippincott-Raven; 1996.
2. Kosinski M, Kujawski SC, Martin R, et al. Health-related quality of life in early rheumatoid arthritis: impact of disease and treatment response. Am J Manag Care 2002;8:231-40.
3. Michaud K, Messer J, Choi HK, Wolfe F. Direct medical costs and their predictors in patients with rheumatoid arthritis: a three-year study of 7,527 patients. Arthritis Rheum 2003;48:2750-62.
4. Yelin E. The costs of rheumatoid arthritis: absolute, incremental, and marginal estimates. J Rheumatol 1996;23 Suppl 44:47-51.
5. Kremer JM, Westhovens R, Leon M, et al. Treatment of rheumatoid arthritis by selective inhibition of T-cell activation with fusion protein CTLA4Ig. N Engl J Med 2003;349:1907-15.
6. Kremer J, Dougados M, Emery P, et al. Treatment of rheumatoid arthritis with the selective costimulation modulator abatacept: Twelve-month results of a phase IIb, double-blind, randomized, placebo-controlled trial. Arthritis Rheum 2005;52:2273-1.
7. Arnett FC, Edworthy SM, Bloch DA, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum 1988;31:315-24.
8. Hochberg MC, Chang RW, Dwosh I, Lindsey S, Pincus T, Wolfe F. The American College of Rheumatology 1991 revised criteria for the classification of global functional status in rheumatoid arthritis. Arthritis Rheum 1992;35:498-502.
9. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992;30:473-83.
10. McHorney CA, Ware JE Jr, Lu JF, Sherbourne CD. The MOS 36-item Short-Form Health Survey (SF-36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. Med Care 1994;32:40-66.
11. McHorney CA, Ware JE Jr, Raczek AE. The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. Med Care 1993;31:247-63.
12. Ware JE Jr, Kosinski M, Gandek B. SF-36® Health Survey: manual and interpretation guide. Lincoln RI: QualityMetric Incorporated; 2000.
13. Ware JE, Kosinski M. SF-36 physical and mental health summary scales: a manual for users of version 1. 2nd ed. Lincoln, RI: QualityMetric Incorporated; 2001.
14. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. J Health Econ 2002;21:271-92.
15. Nunnally JC, Bernstein IH. Psychometric theory. New York: McGraw-Hill, Inc.; 1994.
16. Wyrwich KW, Tierney WM, Wolinsky FD. Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. J Clin Epidemiol 1999;52:861-73.
17. Wyrwich KW, Tierney WM, Wolinsky FD. Using the standard error of measurement to identify important changes on the Asthma Quality of Life Questionnaire. Quality Life Res 2002;11:1-7.
18. Stratford PW, Binkley J, Solomon P, Finch E, Gill C, Moreland J. Defining the minimum level of detectable change for the Roland-Morris questionnaire. Phys Ther 1996;76:359-68.
19. Walters SJ, Brazier J. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. Qual Life Res 2005;14:1523-32.
20. Cohen J. Statistical power for the behavioral sciences. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
21. 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;43:506-14.
22. 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;42:1870-8.
23. Maini SR. Infliximab treatment of rheumatoid arthritis. Rheum Dis Clin North Am 2004;30:329-47; vii.
24. Bang LM, Keating GM. Adalimumab: a review of its use in rheumatoid arthritis. BioDrugs 2004;18:121-39.
25. 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.
26. 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. Arthritis Rheum 2001;44:1984-92.
27. Kalden JR, Scott DL, Smolen JS, et al. Improved functional ability in patients with rheumatoid arthritis: long-term treatment with leflunomide versus sulfasalazine. J Rheumatol 2001;28:1983-91.
28. 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;50:1400-11.
29. 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;22:128-39.
30. Torrance GW, Tugwell P, Amorosi S, Chartash E, Sengupta N. Improvement in health utility among patients with rheumatoid arthritis treated with adalimumab plus methotrexate. Rheumatology Oxford 2004;43:712-8.
31. Strand V. Longer term benefits of treating rheumatoid arthritis: assessment of radiographic damage and physical function in clinical trials. Clin Exp Rheumatol 2004;22 Suppl:S57-S64.
32. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: The remarkable universality of half a standard deviation. Med Care 2003;41:582-92.
33. Walters SJ, Brazier JE. What is the relationship between the minimally important difference and health state utility values? The case of the SF-6D. Health Qual Life Outcomes 2003;1:4.
34. Kosinski M, Schein J, Vallow SM, et al. An observational study of health-related quality of life and pain outcomes in chronic low back pain patients treated with transdermal fentanyl. Curr Med Res Opin 2005;21:849-62.
35. Ware JE Jr, Bayliss MS, Mannocchia M, Davis GL. Health-related quality of life in chronic hepatitis C: Impact of disease and treatment response. The Interventional Therapy Group. Hepatology 1999;30:550-5.