Measuring Preference Weights for American College of Rheumatology Response Criteria for Patients with Rheumatoid Arthritis

CHIUN-FANG CHIOU, MICHAEL WEISMAN, CATHY D. SHERBOURNE, CAROLINA REYES, MICHELLE DYLAN, JOSHUA OFMAN, DANIEL J. WALLACE, WESLEY MIZUTANI, and MARIA E. SUAREZ-ALMAZOR

ABSTRACT. ***Objective.*** To estimate weights for health states comprising American College of Rheumatology (ACR) response and different levels of adverse events associated with rheumatoid arthritis (RA) treatments.

Methods. A survey was mailed to 748 patients with RA from southern California. In addition to several questionnaires commonly used for patients with RA, patients were instructed to evaluate 10 hypothetical health states, in which they could have an ACR response and/or adverse events due to new treatments, with a visual analog scale (VAS). Patients also evaluated their current health with a VAS question and a time tradeoff (TTO) question. Linear extrapolation was used to derive 6 more health states. The Pearson correlation coefficient was used to validate VAS and TTO results.

Results. A total of 487 (65%) patients returned the survey. Among the 10 health states evaluated with VAS directly, the health state in which a patient has ACR70 with no adverse events had the highest VAS weight (0.84), followed by the one having an ACR50 response with no adverse events (0.80). Correlation coefficients ranged from 0.63 for the correlation between VAS and physical component summary to –0.18 between TTO and pain and tender joint count; the correlation coefficients were all statistically significant, indicating there was convergent validity of the VAS and that VAS functioned differently from TTO in how it measured weights.

Conclusion. VAS weights for 16 ACR response health states of patients with RA were derived. These weights could be used for cost-utility analyses of interventions for patients with RA. (J Rheumatol 2005;32:2326–9)

Key Indexing Terms:
RHEUMATOID ARTHRITIS UTILITY TOXICITY
AMERICAN COLLEGE OF RHEUMATOLOGY RESPONSE CRITERIA

To allow the comparison of effectiveness among various medical interventions, an outcome measure — quality-adjusted life year (QALY) — has been recommended by the US panel on cost-effectiveness in health and medicine for use as the effectiveness parameter in any cost-effectiveness analysis[1]. As a comprehensive outcome measure, QALY can simultaneously capture gains from reduced mortality (quantity gains) and reduced morbidity (quality gains). QALY are usually calculated as the product of 2 terms: $\sum(W_i{*}Y_i)$, where Y represents the duration of each health state, W represents the interval-scaled preference weight for a health state, and i represents the specific health states. The preference weights must be measured on or transformed onto an interval scale on which the reference point "death" has a score of 0 and the reference point "optimal health" has a score of 1. The change in QALY due to an intervention can represent the effectiveness of the intervention[1]. Unfortunately, many cost-effectiveness analyses have failed to use QALY as the effectiveness measure due to lack of preference weight for outcome data collected in the clinical trials.

The American College of Rheumatology (ACR) response criteria have commonly been used in recent clinical trials for patients with rheumatoid arthritis (RA) as one of the outcome measures[2-5]. ACR criteria include changes in numbers of swollen joints and tender joints, physician global assess-

From Cerner Health Insights, Beverly Hills; Department of Medicine, Cedars-Sinai Health System, Los Angeles; Department of Rheumatology, Cedars-Sinai Health System, Los Angeles; Health Program, Rand Corporation, Santa Monica; Pharmacoeconomics, Amgen, Thousand Oaks; Talbert Medical Group, Huntington Beach, California; Health Services Research, Baylor College of Medicine, Houston; and Houston Veteran Affairs Medical Center, Houston, Texas, USA.

Supported by a grant from Amgen, Inc., Thousand Oaks, California.

C-F. Chiou, PhD, Cerner Health Insights, Department of Medicine, Cedars-Sinai Health System; M. Weisman, MD, Departments of Medicine and Rheumatology, Cedars-Sinai Health System; C.D. Sherbourne, PhD, Rand Corporation; C. Reyes, PhD; M. Dylan, PhD, Cerner Health Insights; J. Ofman, MD, MSHS, Amgen, Inc.; D.J. Wallace, MD, Departments of Medicine and Rheumatology, Cedars-Sinai Health System; W. Mizutani, MD, Talbert Medical Group; M.E. Suarez-Almazor, MD, PhD, Health Services Research, Baylor College of Medicine, Houston Veteran Affairs Medical Center.

Address reprint requests to M. Dylan, Cerner Health Insights, 9100 Wilshire Blvd., Suite 655E, Beverly Hills, CA 90212, E-mail: mdylan@cerner.com

Accepted for publication June 3, 2005.

ment of pain, C-reactive protein (CRP), erythrocyte sedimentation rate (ESR), and Health Assessment Questionnaire (HAQ) score[6]. An ACR20 response requires a patient to have at least a 20% reduction in the number of swollen and tender joints, and at least a reduction of 20% in 3 of the following 5 indices: physician global assessment of disease, patient global assessment of disease, pain, CRP/ESR, and HAQ score[7]. An ACR50 or an ACR70 response requires a patient to have at least a 50% or 70% reduction. However, a patient meeting ACR response criteria might not have a better quality of life if suffering from adverse events caused by treatments, as compared with a patient who does not meet ACR response criteria but is free from adverse events. Therefore, a comprehensive measure such as QALY would be a more sensitive and valid measure to use when comparing effectiveness across different trials/interventions.

We estimated preference weights for a set of health states comprising ACR20, ACR50, or ACR70 responses and different levels of adverse events associated with treatments for rheumatoid arthritis (RA). These preference weights can be used along with data from existing and/or future clinical trials for comparing the effectiveness of treatments for patients with RA. Our findings may be beneficial for future studies.

MATERIALS AND METHODS

Survey participants. Institutional review board approval was obtained from the Cedars-Sinai Health System (Los Angeles, CA) in November 2002 for administration of our survey. To recruit patients to participate in the survey, an invitation letter was sent to 2234 patients with RA identified by their care providers in 6 rheumatology practices in Los Angeles. A total of 748 patients agreed to participate and were mailed the survey package; from those who were mailed the survey package, 484 completed and returned the surveys. Table 1 lists the sociodemographic and medical characteristics of the study population. The sample included 380 (78.8%) women. Ages ranged from 21 to 91 years (mean 59.4 yrs). Two hundred thirty-nine (49.4%) reported obtaining at least a 2 year college degree. Seventy percent of the participants were non-Hispanic Caucasians. Participants reported having had RA on average for 12.7 years (± 11.2), 14.9 (± 11.7) mean pain and tender joint count and 9.1 (± 9.4) mean swollen joint count out of a possible 48 joints, and mean HAQ score of 1.13 (± 0.75).

Table 1. Patient characteristics.

Total no. participating	484
Mean age (SD), yrs	59.4 (14.1)
Ratio female/male, %	78.5/21.5
Married, %	56.4
Education status	
Completed college courses or higher, %	74.2
Employment status, %	
Working full-time	28.2
Working part-time	8.1
Retired	35
In school	17
Other	11.7
Ethnicity, %	
Asian	6.2
Black, African American	5.8
Latino, Mexican American	12
White, non-Hispanic	70
Other	6
Disease severity	
Mean disease duration (SD), yrs	12.7 (11.2)
Mean no. of painful joints (SD)	14.9 (11.7)
Mean no. of swollen joints (SD)	9.1 (9.4)
Mean HAQ score (SD)	1.13 (0.75)

Survey materials and administration. The survey was designed and developed by a team of health economists, rheumatologists, and psychometricians. The final survey draft was pilot tested among 8 patients with RA. Their comments and suggestions were incorporated into the final version of the survey.

The mailed survey package included the following items: (1) a questionnaire consisting of 28 RA-specific health related quality of life (HRQOL) questions selected by the study steering committee; (2) a self-assessment of their current health state on a 100 point visual analog scale (VAS); (3) a pictorial mannequin for the number of swollen joints and the number of painful and tender joints[8]; (4) a set of hypothetical health states related to ACR response and adverse events caused by treatments for RA evaluated on a 100 point VAS on which 0 represents "death" and 100 represents "perfect health"; (5) a time tradeoff (TTO) question for the health state in which a patient has no improvement and no adverse events; (6) HAQ; (7) the Medical Outcomes Study Short-Form 36 (SF-36) questionnaire[9]; and (8) sociodemographic and medical questions. A reminder telephone call was made to those who did not return the first mailing within 2 weeks. Patients who returned the survey were compensated for participating ($25).

Responses to the set of VAS questions on the hypothetical health states were used to estimate the weights for ACR20 related health states directly. In these questions, respondents were asked to use the VAS technique to evaluate hypothetical health states in which they would take a new treatment for RA for 2 months. Respondents were asked to rate the state of their health after imagining taking a medication that improved their joint pain and swelling and decreased the amount of RA pain by at least 20% as well as improving their overall health by 20%; but at the same time developing adverse effects that would require treatment by a doctor. In each health state, the respondent would either achieve ACR response or not and would have different levels of adverse events associated with it. Since ACR response is measured relative to a patient's health state at baseline, we instructed patients to complete the 28 RA-specific HRQOL questions and joint counts first. Therefore, patients would have a good understanding of their current health states prior to answering the ACR response related questions and thus would provide more valid evaluations. To limit respondents' burden, weights for most ACR50 and ACR70 related health states were extrapolated based on these data. For example, the weight for a health state in which a patient has an ACR50 response with severe adverse events was derived by adding the "difference between the weight for a health state in which a patient has ACR50 response with no adverse events and the one for ACR20 response with no adverse events" to the "weight for the health state in which a patient has ACR20 response with severe adverse events." It is worth noting that adverse events included in the health states were categorized as none, mild (injection site reaction, headache, rhinitis, dizziness, asthenia, abdomen pain, rash, and dyspepsia), moderate (upper respiratory infection, pharyngitis, respiratory disorder, sinusitis), or severe (gastrointestinal bleed, sepsis, pneumonia, and any diseases requiring hospitalization) by RA experts.

Analysis. To test whether the VAS and TTO had validity, Pearson correlation coefficients (PCC) were estimated for VAS and for TTO scores for the general health with the SF-36 Mental Component Summary (MCS) score, SF-36 Physical Component Summary (PCS) score, HAQ score, and joint counts, respectively. All statistical analyses were performed using SAS version 8.2 for Windows (SAS Institute, Cary, NC, USA).

RESULTS

The mean VAS and TTO scores were 0.647 (± 0.179) and

0.849 (± 0.263), respectively; this result is consistent with a previous study in which the TTO score was greater than the VAS score[10]. The mean SF-36 PCS and MCS scores were 36.6 (± 9.4) and 43.7 (± 10.8), respectively; the mean PCS score was between that of people with minor medical problems (i.e., 46.37) and people with serious medical problems (i.e., 36.27). The mean MCS score was higher than that of people with psychiatric problems (i.e., 37.62), but lower than that of people with minor medical problems (i.e., 54.29)[11].

Table 2 lists VAS weights for all health states of interest. Among the 10 health states evaluated with VAS directly, the health state in which a patient has ACR70 with no adverse events had the highest VAS weight (0.84), followed by the one having an ACR50 response with no adverse events (0.80) and the one having an ACR20 response with no adverse events (0.68). In every level of the adverse events associated health states, health states with no improvement had lower weights than those having an ACR response.

Table 3 lists all PCC, ranging from 0.63 for the correlation between VAS and PCS to –0.18 between TTO and pain and tender joint count. Although all PCC were statistically significant and in the right directions as expected, those of VAS and other measures were consistently higher than those of TTO and the same measures. It indicated that the VAS and TTO techniques worked differently when measuring the same health state in our study.

DISCUSSION

With limited resources and ever-increasing demand for health services, it is especially important to have a comprehensive measure of effectiveness to facilitate comparison across different interventions. We derived a set of weights for health states of patients with RA; these health states were defined with ACR response and the existence of adverse events caused by treatments for RA — the 2 most commonly used outcome measures of recent clinical trials for RA. The weights can be used to estimate QALY of patients in each intervention and to facilitate comparisons across different interventions used for patients with RA. For example, among patients taking treatment A for 8 weeks, 20% had an ACR70 response, 30% had an ACR50 response, 40% had an ACR20 response, and the remainder had no improvement; 40% of them had mild adverse events, 30% had moderate adverse events, and 20% had severe adverse events. Assuming that all other aspects were equal and that 100% of patients were in the health state with no improvement and no adverse events, the QALY gained due to treatment A over a year would be 0.1064. If the same patient population can gain more than 0.1064 QALY (e.g., 0.21) with treatment B and if treatment B is cheaper than treatment A (e.g., $1000 per year), then treatment B would be more cost-effective than treatment A, with an incremental cost-effectiveness ratio equal to $10,000 per QALY. With the same approach, the aforementioned QALY can be used to compare those of patients with different diseases to enable health decision-makers to make more optimal allocations for the limited resources among interventions for different diseases.

The weight for the health state with ACR70 response was greater than that with ACR50 response; the one with ACR50 response was greater than the one with ACR20 response; either was consistent with medical interpretation of the ACR response criteria. With no exception, the weight for the health state with no adverse events or less severe adverse events was also greater than for those with severe adverse events. However, the direct elicitation technique has been shown to be more likely to collect upwardly skewed data[12]. In addition, the description used for health states in this study did not include all components of ACR response criteria due to our concern that it would be difficult for patients to relate data such as CRP, ESR, and HAQ scores to their

Table 2. Derived and directly estimated VAS weights for health states*.

Adverse Events	No Improvement	ACR20 Response	ACR50 Response	ACR70 Response
None	0.53	0.68	0.80	0.84
Mild	0.49	0.64	0.76†	0.80†
Moderate	0.46	0.58	0.70†	0.74†
Severe	0.34	0.41	0.53†	0.57†

* VAS ranges from 0 to 1. † Derived based on directly estimated VAS weights.

Table 3. Pearson correlation coefficients of VAS and TTO vs other measures*.

	PCS	MCS	HAQ	Pain and Tender Joints	Swollen Joints
VAS score	0.63	0.65	–0.59	–0.39	–0.35
TTO score	0.26	0.24	–0.29	–0.18	–0.20

* All p values < 0.00. PCS: physical component summary, MCS: mental component summary, HAQ: Health Assessment Questionnaire, VAS: Visual analog scale, TTO: time tradeoff.

current health states. Further, many adverse events can vary greatly in severity, and patients might not be able to rate health states related to adverse events they had not experienced. Finally, due to the concern over patients' cognitive burden, the weights of most of the health states with ACR50 or ACR70 responses were not directly elicited from patients. A patient with ACR70 response might not be exactly as critical of adverse events as a patient with ACR50 or ACR20 response. This issue warrants further studies.

It is not clear whether either of the aforementioned factors would compromise the validity of the study results. Fortunately, convergent validity of the VAS direct elicitation technique seemed to exist; it was supported by all Pearson correlation coefficients. Nonetheless, it might be worth using the relative risk attitude equation[13] or the multiattribute utility function technique in future clinical trials to collect another set of preference weights.

Since ACR response is a relative measure (i.e., relative to the health state of a patient at baseline), users of the weights derived in this study should be aware of the characteristics of our study population. Our patients were a convenience sample recruited from 2 counties in Southern California. While the gender distribution of the sample (predominantly female) was roughly the same as the typical RA population[14], the sample was skewed toward people with a relatively high socioeconomic status who classified themselves as non-Hispanic whites. In addition, few of the patients (3.5%) reported having severe physical disability based on the HAQ. The survey materials received by the study population included a 28 item questionnaire, as well as the SF-36, the HAQ, and several other types of questionnaires. The difference in baseline characteristics between future studies and our study should not limit the validity of the derived weights when applied to estimate QALY of different interventions within the same population and to make a comparison. Whether differences across populations may affect these weights remains to be determined in future studies.

This is the first study to assign patient-derived weights to ACR response criteria. These weights can be used to conduct cost-utility studies of interventions for RA.

REFERENCES

1. Identifying and valuing outcomes. In: Gold M, Siegal J, Russel L, Weinstein M, editors. Cost-effectiveness in health and medicine. New York: Oxford University Press; 1996:82-134.
2. Weinblatt ME, Kremer JM, Bankhurst AD, et al. A trial of etanercept, a recombinant tumor necrosis factor receptor:Fc fusion protein, in patients with rheumatoid arthritis receiving methotrexate. N Engl J Med 1999;340:253-9.
3. Maini R, St. Clair EW, Breedveld F, et al. Infliximab (chimeric anti-tumour necrosis factor alpha monoclonal antibody) versus placebo in rheumatoid arthritis patients receiving concomitant methotrexate: a randomised phase III trial. ATTRACT Study Group. Lancet 1999;354:1932-9.
4. Genovese MC, Bathon JM, Martin RW, et al. Etanercept versus methotrexate in patients with early rheumatoid arthritis: two-year radiographic and clinical outcomes. Arthritis Rheum 2002;46:1443-50.
5. Cohen S, Hurd E, Cush J, et al. Treatment of rheumatoid arthritis with anakinra, a recombinant human interleukin-1 receptor antagonist, in combination with methotrexate: results of a twenty-four-week, multicenter, randomized, double-blind, placebo-controlled trial. Arthritis Rheum 2002;46:614-24.
6. Fries JF, Spitz PW, Young DY. The dimensions of health outcomes: the Health Assessment Questionnaire, disability and pain scales. J Rheumatol 1982;9:789-93.
7. Felson DT, Anderson JJ, Lange ML, Wells G, LaValley MP. Should improvement in rheumatoid arthritis clinical trials be defined as fifty percent or seventy percent improvement in core set measures, rather than twenty percent? Arthritis Rheum 1998;41:1564-70.
8. Wong AL, Wong WK, Harker J, et al. Patient self-report tender and swollen joint counts in early rheumatoid arthritis. Western Consortium of Practicing Rheumatologists. J Rheumatol 1999;26:2551-61.
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. Bleichrodt H, Johannesson M. Standard gamble, time trade-off and rating scale: experimental results on the ranking properties of QALYs. J Health Econ 1997;16:155-75.
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. Donald SC, Unutzer J, Schoenbaum M, et al. Can utility-weighted health-related quality-of-life estimates capture health effects of quality improvement for depression? Med Care 2001;39:1246-59.
13. Torrance GW, Feeny DH, Furlong WJ, Barr RD, Zhang Y, Wang Q. Multiattribute utility function for a comprehensive health status classification system. Health Utilities Index Mark 2. Med Care 1996;34:702-22.
14. Lawrence RC, Helmick CG, Arnett FC, et al. Estimates of the prevalence of arthritis and selected musculoskeletal disorders in the United States. Arthritis Rheum 1998;41:778-99.