Preliminary Evaluation of a Visual Analog Function Scale for Use in Rheumatoid Arthritis

FREDERICK WOLFE, KALEB MICHAUD, and THEODORE PINCUS

ABSTRACT. ***Objective.*** Key outcomes in rheumatoid arthritis (RA) are evaluated with multi-item ratings scales such as the Health Assessment Questionnaire (HAQ) and visual analog scales (VAS) such as pain and patient and physician global. As VAS scales are easy to use and particularly effective in research and patient care, we studied the characteristics, association, and psychometric properties of a VAS function scale (VAS-F) to determine if it could be used in RA studies and clinical practice.

Methods. A total of 394 patients with RA completed the HAQ, the HAQ-II, and a VAS functional scale. In addition, they completed standard assessments of pain, global, fatigue, sleep problems, joint count, and the Medical Outcome Study Short-Form 36 (SF-36) physical component summary score (PCS) and vitality and total pain scores.

Results. The HAQ-II was correlated with VAS-F at 0.76, but distributional characteristics of the HAQ and VAS-F differed, as the VAS-F scale results contained more higher scores as well as more lower scores compared with the HAQ-II and HAQ. Kendall's tau concordance analyses indicated that VAS scales were more concordant with other VAS than with non-VAS scales. Concordance of VAS-F was greatest with VAS global and was similar overall with VAS pain, sleep disturbance, fatigue, and quality of life. By contrast, the PCS, a multi-item scale, was more concordant with HAQ-II and HAQ. There was little to no difference between the VAS-F and the 2 HAQ with regard to concordance with the multi-item joint count, SF-36 vitality, and SF-36 total pain.

Conclusion. The distribution differences between HAQ and HAQ-II and the VAS-F suggest that patients do not see minor limitations as problematic, but rate major limitations as being particularly limiting and worthy of high ratings. A VAS functional scale represents a patient-weighted functional assessment in which additional interpretation is given to the meaning of the limitations by the patient. VAS-F scales may be suitable for use in the clinic and in research. However, studies to assess sensitivity to change are required to determine the appropriate role of this scale. (J Rheumatol 2005;32:1261–6)

Key Indexing Terms:
VISUAL ANALOG SCALE HEALTH ASSESSMENT QUESTIONNAIRE
HEALTH ASSESSMENT QUESTIONNAIRE-II FUNCTIONAL ASSESSMENT
RHEUMATOID ARTHRITIS

Two kinds of measurement scales dominate rheumatoid arthritis (RA) evaluations. The first, multi-item rating scales, consist of a series of questions — often called items — that are categorical and represent, for example, concepts such as mild, moderate, and severe. The items of the scale are summed or averaged to produce a resultant score. A good example of this type of scale is the Health Assessment Questionnaire (HAQ)[1] or the Health Assessment Questionnaire II (HAQ-II)[2], in which the results of the 8 categories or 10 questions of the respective scales are averaged. The second kind of scale is a summary scale that is represented only by a single item. This scale may take the form of a visual analog scale (VAS) or a numeric rating scale (NRS). The VAS is represented by a single line (usually 10 or 15 cm) anchored at one end with a descriptor such as "no problem or difficulty" and at the other end with a descriptor such as "severe problems or difficulties." Patients place a mark on the line to indicate how they evaluate their condition. The distance between the 0 point and the mark on the line is calculated and a value between 0 and 10 is obtained. NRS may hold as few as 3 categories and as many as 20 or more. When there are many categories NRS and VAS have similar, although not the same, properties.

In RA the most commonly used scales are the HAQ and modified HAQ (MHAQ)[3], the Medical Outcome Study Short-Form 36 (SF-36)[4] multi-item rating scales, and VAS for pain, patient global, and physician global. Assuming the items of a multi-item scale are well chosen and appropriate, a multi-item rating scale, by virtue of its many questions, will be more accurate and reliable than a single-item VAS.

From the Arthritis Research Center Foundation, University of Kansas School of Medicine, Wichita, Kansas; and Division of Rheumatology, Vanderbilt University, Nashville, Tennessee, USA.

Supported by a grant from Abbott Laboratories, Abbott Park, IL, USA.

F. Wolfe, MD, Arthritis Research Center Foundation, University of Kansas School of Medicine; K. Michaud, MS, Arthritis Research Center Foundation; T. Pincus, MD, Division of Rheumatology, Vanderbilt University.

Address correspondence to Dr. F. Wolfe, National Data Bank for Rheumatic Diseases, Arthritis Research Center Foundation, 1035 N. Emporia, Suite 230, Wichita, KS 67214.
E-mail: fwolfe@arthritis-research.org

Accepted for publication February 21, 2005.

However, this complexity comes at a price of patient and physician burden. VAS scales are easier to use and score, but are thought to be less accurate and reliable. Despite this limitation, VAS scales are almost universally used for the evaluation of pain and physician and patient global in randomized clinical trials (RCT) and clinical care.

Despite the theoretical objection to VAS scales, they often outperform rating scales in their ability to distinguish treatment effects in RCT. It has long been believed that the reason for this is that pain or global is a concept that is more sensitive to the effects of treatment than is function. However, there is another possible explanation, that VAS scales more accurately convey what patients are feeling than do multi-item rating scales. If this is the case, it is possible that a VAS function scale might have a role in the evaluation of RA.

We evaluated the psychometric and associative properties of a VAS function scale compared with the HAQ-II and HAQ, and compared the concordance of the function scale and VAS scales in general with multi-item rating scales.

MATERIALS AND METHODS

Patient sample. Patients in this study were participants in the National Data Bank for Rheumatic Diseases (NDB) longitudinal study of RA outcomes. Patients are recruited from the practices of United States rheumatologists[5-7], and are followed with semiannual questionnaires. This report concerns the status of 394 randomly selected patients with RA who completed an internet based survey questionnaire in the first half of 2004.

Demographic and disease status variables. NDB participants completed semiannual, detailed 28 page questionnaires about all aspects of their illness. At each assessment, demographic variables were recorded including sex, age, ethnic origin, education level, current marital status, and medical history. Functional assessment measures include the Stanford HAQ functional disability index[1], the modified HAQ[3], the HAQ-II, a shortened, modified version of the HAQ with similar scaling but superior psychometric properties[2], and the SF-36, from which the physical component summary score (PCS) was calculated[4,8]. We also utilized the SF-36 total pain and vitality dimension questionnaire. VAS scales included those for gastrointestinal (GI) problems, fatigue, sleep, pain, and global severity, as described. A self-report joint count was obtained using the RA Disease Activity Index (RADAI) scales[9]. The VAS-F functional scale was anchored at one end with "No functional limitations" and the other end with "Severe functional limitations."

Statistical analyses. Kendall's tau-a and associated confidence limits were calculated using the Somers-D package[10]. Tau-a has a simple interpretation, the percentage agreement between function scales and other clinical variables. By contrast, Pearson correlation coefficients do not have a simple linear interpretation. In addition, the nonparametric tau-a is robust against extreme values. Tau-a is further interpreted in the Results section below. Smoothed lines of Figures 2 and 3 were produced by kernel-weighted local polynomial smooths using the Epanechnikov kernel. Correlation analyses and other analyses were performed using Stata v. 8.0[11]. Rasch item estimates for the HAQ-II were calculated using Winsteps v. 3.53[12].

RESULTS

The mean (SD) age and disease duration of the 394 participants were 56.3 (12.1) years and 15.5 (10.2) years, respectively; 22.6% of participants were male (Table 1). The 3 function scales had the same mean (0.9). However, the medians were different, being 0.8 for HAQ and HAQ-II, but 0.6 for the VAS functional scale. In addition, the standard deviation was greatest for the VAS (0.8), intermediate for the HAQ (0.7), and smallest for the HAQ-II (0.6).

Table 1. Demographic and clinical characteristics of 394 patients with RA.

Variable	Mean	SD	Median
Age, yrs	56.2	12.1	56.2
Sex, % male	22.6		
Disease duration, yrs	15.5	10.0	13.6
HAQ (0–3)	0.9	0.7	0.8
HAQ-II (0–3)	0.9	0.6	0.8
VAS function scale (0–10)	0.9	0.8	0.6
Pain (0–10)	3.8	2.8	3.0
Global severity (0–10)	3.0	2.3	2.5
Joint count (self-report) (0–16)	7.4	4.5	7.0
Sleep disturbance (0–10)	3.5	3.0	3.0
Fatigue (0–10)	4.2	2.9	4.0
VAS QOL scale (0–100)	86.3	4.2	85.0
Physical component score (SF-36)	34.8	10.6	35.6
GI severity (0–10)	1.8	2.2	1.0
SF-36 total pain (0–100)	39.8	14.9	41.0
SF-36 vitality scale (0–100)	45.6	23.8	45.0

HAQ: Health Assessment Questionnaire, VAS: visual analog scale, QOL: quality of life, GI: gastrointestinal.

Figure 1 displays HAQ-II values and VAS-F values at common quantiles, and a moderate amount of random noise has been added to make the data points more visible. If the values were the same they would all lie on the line of agreement. However, shifting occurs, such that above the mean (~0.9) values are greater for the VAS-F than for the HAQ-II. In addition, for values below the mean VAS-F, scores are lower than are HAQ-II scores. This is also seen in Figures 2 (HAQ-II) and 3 (HAQ), particularly at high levels, where values of 3 on the VAS-F scale correspond to values of approximately 2 on the HAQ-II and HAQ scales. In addition, the mean value of HAQ-II and HAQ at the intercept (where VAS-F is 0) are approximately 0.3. These data indicate that patients rating their function on the VAS-F scale indicate higher scores at higher levels of dysfunction and lower scores at lower levels of dysfunction compared with HAQ-II and HAQ scores.

Although the HAQ has been shown to be a relatively linear scale by Rasch analyses, it still does not correspond to a fully linear scale. To understand the relationship between the VAS-F and a fully linear scale, we obtained Rasch difficulty scores and analyzed them with a quantile-quantile plot (Figure 4). Each unit on the y-axis represents 1 logit. As with Figure 1, levels above the HAQ-II mean show a shifting from the line of identity. For VAS-F scores from 0.25 to 3, the curves of Figure 1 and Figure 4 are essentially identical, indicating the overall linearity of the HAQ-II scale. However, there is a significant difference in the graphs at levels below 0.25, which corresponds to the intercept of ~0.3 in Figures 2 and 3. There is a 10-logit difference

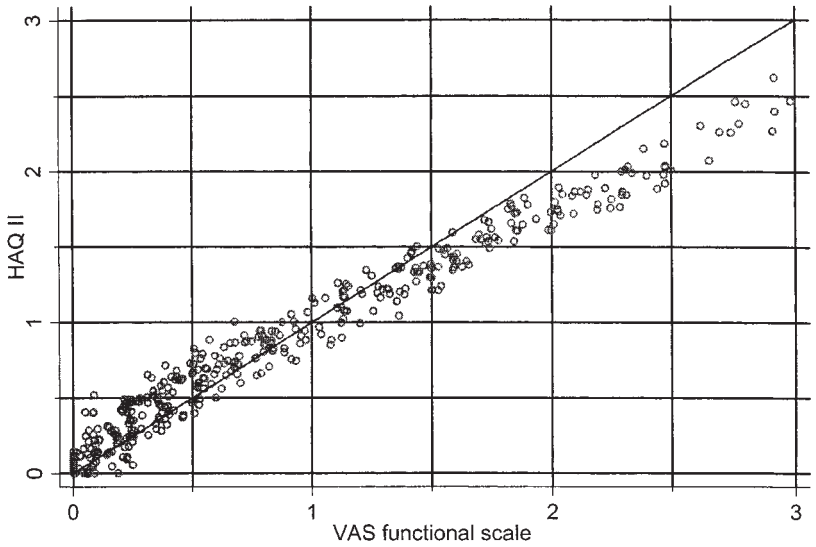

Figure 1. Plot of quantiles of HAQ-II against VAS-F. At quantiles above the median, VAS-F scores are greater than HAQ-II score. At quantiles below the median, VAS-F scores are lower than HAQ-II score. A moderate amount of random noise has been added to make the data points more visible.

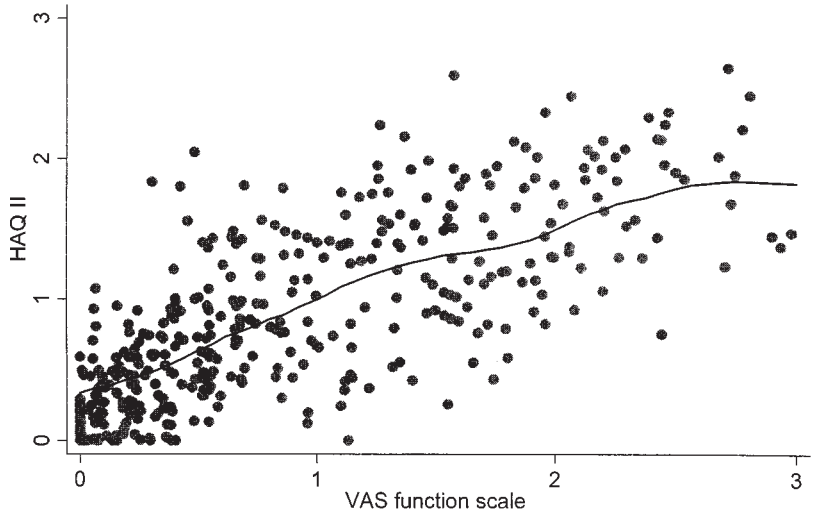

Figure 2. The HAQ-II versus the VAS-F scale. The line is generally linear, but the intercept crosses the y-axis at ~0.3.

between a VAS-F score of 0.25 and 0, and a 1 to 3-logit difference between VAS-F scores of 0 and 0.125. These data indicate that zeroes on the HAQ scales do not necessarily mean no functional limitation. By contrast, a VAS-F score comes closer to the no-limitation interpretation. Clinically, this is an important point. These data show that a HAQ score of 0 is often associated with functional disability that cannot be measured by the HAQ (but can be measured by a VAS).

As the VAS-F and HAQ scales differed in their distributional characteristics, we investigated whether the 2 types of scales might have dissimilar clinical effects. Table 2 displays levels of agreement between the 3 functional scales and RA clinical variables using Kendall's tau-a. The coefficients in columns 2–4 can be interpreted as percentage agreement between each functional variable and the RA variables of column 1. For example, VAS global and VAS function are 60.4% more likely to agree than to disagree. By contrast, VAS global and HAQ-II are only 41.5% more likely to agree than to disagree. The half-difference between tau-a for VAS global-VAS function and tau-a for VAS global–HAQ-II is equal to the difference between 2 probabilities, namely the probability that VAS global is concordant with VAS function and not with HAQ-II, and the probability that VAS global is concordant with HAQ-II and not with

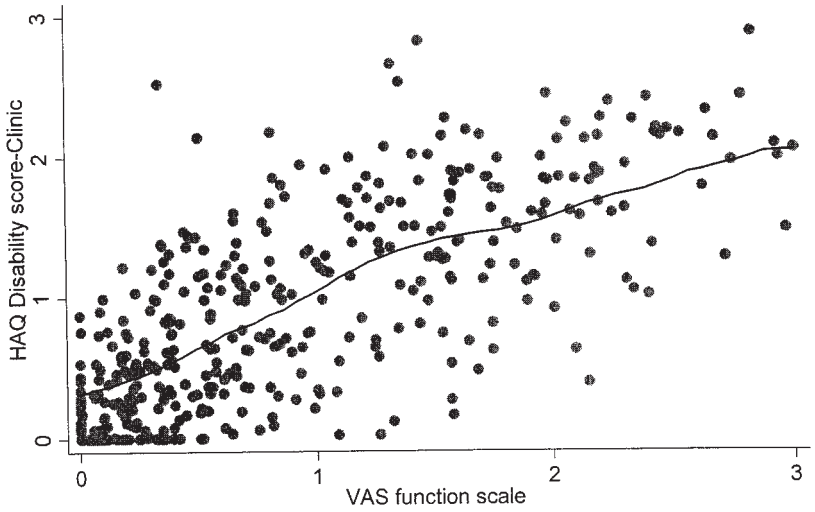

Figure 3. The HAQ versus the VAS-F scale. The line is generally linear, but the intercept crosses the y-axis at ~0.3.

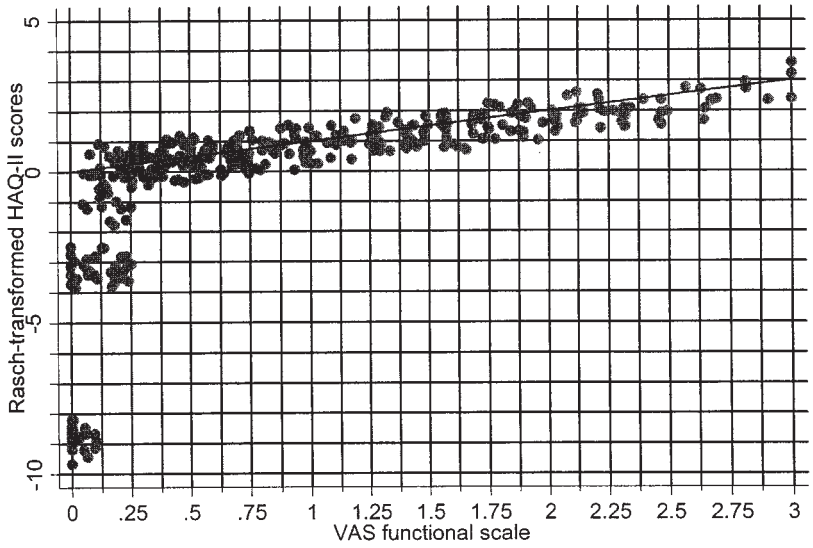

Figure 4. Plot of quantiles of Rasch-derived item difficulties of HAQ-II against VAS-F. Y-scale units are in logits. At quantiles above the median VAS-F, scores are greater than HAQ-II logit score. At quantiles below the median but above 0.25 on the VAS-F scale, VAS-F scores are lower than HAQ-II scores. These data indicate that zeroes on the HAQ scales do not necessarily mean no functional limitation. By contrast, a VAS-F score comes closer to the no-limitation interpretation.

VAS. The half-differences, then, are an overall measure of concordance. For example, VAS global is 7%–12% more likely to be concordant with VAS function (but not with HAQ-II) than with HAQ-II (but not with VAS function).

The half-difference results show that the VAS scales are more concordant with other VAS scales than with non-VAS variables. Concordance of VAS function is greatest with VAS global and similar overall with VAS pain, sleep disturbance, fatigue, and quality of life (QOL). By contrast, PCS, a multi-item scale, is more concordant with HAQ-II and HAQ. There is little to no difference between the VAS function and the 2 HAQ with regard to concordance with joint count, SF-36 vitality, and SF-36 total pain, all these scales being multi-item rating scales, not VAS.

Table 3 shows the strength of the relationship between the study variables using conventional Pearson correlation analysis.

DISCUSSION

Although mean values are similar for VAS-F and HAQ scores, the distribution of VAS scores differs substantially from the distribution of HAQ scores. Compared with the 2 HAQ, the VAS-F scale results contain more higher scores as well as more lower scores (Figures 1–4). The likely interpretation of this finding is that patients do not see minor limitations as problematic, but rate major limitations as being particularly limiting and worthy of high ratings. Such a VAS functional scale represents a patient-weighted functional

Table 2. Correlation between the 3 function scales and other clinical variables.

Variables	VAS function	HAQ-II	HAQ
VAS pain	0.774	0.635	0.635
VAS global	0.759	0.565	0.550
PCS	–0.731	–0.825	–0.763
VAS fatigue	0.672	0.545	0.499
VAS sleep	0.605	0.461	0.418
SF-36 vitality	–0.575	–0.540	–0.479
Joint count (self-report)	0.551	0.544	0.560
VAS QOL	0.515	0.436	0.407
SF-36 total pain score	–0.461	–0.421	–0.412
VAS GI scale	0.356	0.337	0.312

PCS: SF-36 physical component summary score.

assessment in which additional interpretation is given to the meaning of the limitations by the patient. By contrast, multi-item scales such as the HAQ-II, HAQ, and Western Ontario and McMaster University Osteoarthritis Index (WOMAC) can provide a precise measure of functional ability, as each item/category is associated with a particular level of difficulty. Such difficulty levels can be demonstrated by Rasch analysis.

In addition to precisely definable levels of difficulty, multi-item rating scales have smaller standard deviations, which in RCT and observational studies can mean smaller sample sizes. For example, a 0.25 difference (from 1.75 to 1.50) in means using the HAQ-II requires a sample size of 122 subjects using the standard deviations noted in this study. By contrast, the VAS-F would require 216 subjects. However, smaller standard deviations do not mean that multi-item rating scales will be more sensitive to change than VAS scales. In RCT, VAS scales for pain and patient global are often among the measures most sensitive to change[13,14]. To our knowledge there have been no RA studies that have used a VAS functional scale, so its comparative value is not known. If a VAS functional scale were shown to have relative efficiency similar to the HAQ, this would allow for reduced patient burden, something that might facilitate the use of assessment questionnaires in the clinic. The strong correlation of the VAS-F with patient pain and global shown in Table 2, and the greater concordance of the VAS-F than the 2 HAQ with patient global and pain, suggest the possibility that VAS-F would be a useful and sensitive scale.

Functional questionnaires have many roles. One role is to measure functional status. In that respect, a multi-item rating scale will be superior to a VAS scale. Where precise measurements of functional status may be less important, such as in the measurement of changes in RCT or the clinic, a VAS scale may be particularly useful. Although this study shows psychometric and theoretical reasons why a VAS-F scale could work, it remains for treatment studies to evaluate the full role of the VAS-F scale. We present the data from our study to stimulate use of the VAS-F scale in such studies. One RCT is now in progress that should provide data on the relative efficiency of the 2 HAQ compared with the VAS-F.

Finally, the VAS-F may have other values. In comparison to the Health Assessment Questionnaires, the VAS can be scored instantaneously, and it has an intuitive meaning. By contrast, very few people have a full sense of what a HAQ score means and how it should be interpreted. It is possible that a simple VAS scale will induce more physicians to measure function in the clinic. Even so, additional studies will be required to understand if the gain in simplicity and ease of use will balance the loss of precision and accuracy that comes with a multi-item functional scale.

REFERENCES

1. Fries JF, Spitz PW, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. Arthritis Rheum 1980;23:137-45.
2. Wolfe F, Michaud K, Pincus T. HAQ-II: Development and validation of a revised version of the Health Assessment Questionnaire (HAQ). Arthritis Rheum 2004;50:3296-305.
3. Pincus T, Summey JA, Soraci SA Jr, Wallston KA, Hummon NP. Assessment of patient satisfaction in activities of daily living using

Table 3. Levels of agreement between functional assessment variables and other clinical variables.

Variable	VAS Function	HAQ II	HAQ	HAQ-II Half-Difference (95% CI)	HAQ Half-Difference (95% CI)
VAS sclaes					
VAS global	0.604	0.415	0.403	0.094 (0.073–0.116)	0.101 (0.078–0.123)
VAS pain	0.583	0.465	0.456	0.059 (0.038–0.080)	0.063 (0.041–0.087)
VAS sleep	0.429	0.324	0.291	0.058 (0.037–0.079)	0.078 (0.056–0.101)
VAS fatigue	0.501	0.386	0.346	0.052 (0.030–0.075)	0.069 (0.046–0.091)
VAS QOL	0.446	0.363	0.334	0.042 (0.019–0.064)	0.056 (0.033–0.080)
VAS GI scale	0.288	0.250	0.231	0.019 (–0.002–0.040)	0.028 (0.005–0.052)
Non-VAS scales					
Joint count	0.377	0.376	0.393	0.000 (–0.021–0.022)	–0.008 (–0.031–0.015)
SF-36 vitality	0.419	0.394	0.341	–0.012 (–0.034–0.009)	–0.039 (–0.062–(–)0.017)
SF-36 total pain	0.362	0.337	0.320	–0.012 (–0.034–0.004)	–0.021 (–0.044–0.022)
PCS	0.553	0.626	0.560	0.037 (0.018–0.057)	0.004 (–0.017–0.048)

a modified Stanford Health Assessment Questionnaire. Arthritis Rheum 1983;26:1346-53.
4. Ware JE, Sherbourne CD. The MOS 36-Item Short-Form Health Survey (SF-36). 1. Conceptual framework and item selection. Med Care 1992;30:473-83.
5. Wolfe F, Anderson J, Burke TA, Arguelles LM, Pettitt D. Gastroprotective therapy and risk of gastrointestinal ulcers: risk reduction by COX-2 therapy. J Rheumatol 2002;29:467-73.
6. Wolfe F, Flowers N, Burke TA, Arguelles LM, Pettitt D. Increase in lifetime adverse drug reactions, service utilization, and disease severity among patients who will start COX-2 specific inhibitors: quantitative assessment of channeling bias and confounding by indication in 6689 patients with rheumatoid arthritis and osteoarthritis. J Rheumatol 2002;29:1015-22.
7. Michaud K, Messer J, Choi HK, Wolfe F. Direct medical costs and their predictors in persons with rheumatoid arthritis: a 3 year study of 7,527 patients. Arthritis Rheum 2003;48:2750-62.
8. 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.
9. Stucki G, Liang MH, Stucki S, Bruhlmann P, Michel BA. A self-administered rheumatoid arthritis disease activity index (RADAI) for epidemiologic research. Psychometric properties and correlation with parameters of disease activity. Arthritis Rheum 1995;38:795-8.
10. Newson R. Parameters beyond "nonparametric" statistics: Kendall's tau, Somer's D and median differences. Stata Journal 2002;2:45-64.
11. Stata Corporation. Stata statistical software: release 8.2. College Station, TX: Stata Corp.; 2003.
12. Linacre JM. Facets Rasch measurement computer program. Chicago: Winsteps; 2004. Avaialble from: Winsteps.com. Accessed March 18, 2005.
13. 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. [Erratum: Arthritis Rheum 2000;43:1345].
14. Anderson JJ, Bolognese JA, Felson DT. Comparison of rheumatoid arthritis clinical trial outcome measures: a simulation study. Arthritis Rheum 2003;48:3031-8.