

The Minimally Important Difference for the Fatigue Visual Analog Scale in Patients with Rheumatoid Arthritis Followed in an Academic Clinical Practice

DINESH KHANNA, JANET E. POPE, PUJA P. KHANNA, MICHELLE MALONEY, NOOSHIN SAMEDI, DEBBIE NORRIE, GILLIAN OUIMET, and RON D. HAYS

ABSTRACT. Objective. To estimate the minimally important difference (MID) for a fatigue visual analog scale (VAS) using patient-reported anchors (fatigue, pain, and overall health).

Methods. Patients with rheumatoid arthritis (RA; n = 307) had 2 clinic visits at a median of 5.9 months apart. They completed a fatigue VAS (0–10 scale) and the retrospective anchor items, “How would you describe your overall fatigue/pain/overall health since the last visit?” with response options: Much worsened, Somewhat worsened, Same, Somewhat better, or Much better. The fatigue anchor was used for primary analysis and the pain/overall health anchors for sensitivity analyses. The minimally changed group was defined by those reporting they were somewhat better or somewhat worsened.

Results. The mean [standard deviation (SD)] age was 59.4 (13.2) years, disease duration was 14.1 (11.5) years, and 83% of patients were women. The baseline mean (SD) Health Assessment Questionnaire–Disability Index score was 0.84 (0.75). The baseline fatigue VAS score was 4.2 (2.9) and at followup was 4.3 (2.8) [mean change of –0.07 (2.5); p = not significant]. The fatigue change score (0–10 scale) for Somewhat better and Somewhat worsened for the fatigue anchor averaged –1.12 and 1.26, respectively. Using the pain anchor, the fatigue change score for Somewhat better and Somewhat worsened averaged –0.87 and 1.13; and using the global anchor, the fatigue change score for Somewhat better and Somewhat worsened averaged –0.82 and 1.17, respectively. Effect size estimates using 3 anchors were small for the Somewhat better (range 0.27–0.39) and Somewhat worsened (0.40–0.44) groups, but larger than for the no-change group (0.03–0.08).

Conclusion. The MID for fatigue VAS is between –0.82 for –1.12 for improvement and is 1.13 to 1.26 for worsening on a 0–10 scale in a large RA clinical practice, and is similar to that seen in RA clinical trials. This information can aid in interpreting fatigue VAS in day-to-day care in clinical practice. (First Release Nov 1 2008; J Rheumatol 2008;35:2339–43; doi:10.3899/jrheum.080375)

Key Indexing Terms:

MINIMAL CLINICALLY IMPORTANT DIFFERENCES
FATIGUE
RELIABLE CHANGE INDEX

RHEUMATOID ARTHRITIS
VISUAL ANALOG SCALE
CLINICAL PRACTICE

Fatigue is very common in patients with rheumatoid arthritis (RA) and is associated with poor day-to-day functioning and overall sense of well-being¹. Fatigue is related to the

RA disease activity and other non-RA factors (poor sleep, stress, etc). The prevalence of fatigue in RA, defined as a rating of ≥ 2 on a 0 to 10 visual analog scale (VAS), is

From the Division of Rheumatology, Department of Medicine, David Geffen School of Medicine; Department of Health Services, School of Public Health, University of California at Los Angeles, Los Angeles, California, USA; Division of Rheumatology, St. Joseph's Health Care, University of Western Ontario, London, Ontario, Canada; and Rand Corporation, Santa Monica, California, USA.

Dr. D. Khanna was supported by a National Institutes of Health Award (NIAMS K23 AR053858-01A1). Dr. P.P. Khanna was supported by a National Institutes of Health Award (T32 AR 053463). Dr. Hays was supported by the UCLA Center for Health Improvement in Minority Elderly/Resource Centers for Minority Aging Research, NIH/NIA/NCMHHD, under Grant 2P30-AG-021684 and a grant from the National Institute on Aging (AG 020679-01). Pfizer Inc. and Bristol-Meyer Squibs Inc. provided unrestricted grants for portions of the study. D. Khanna, MD, MS, Assistant Professor of Medicine, Division of Rheumatology, Department of Medicine, David Geffen School of Medicine, UCLA, and Department of Health Services, School of Public

Health, UCLA; J.E. Pope, MD, MPH, FRCPC, Professor of Medicine, Division of Rheumatology, St. Joseph's Health Care and University of Western Ontario; P.P. Khanna, MD, MPH, Fellow in Rheumatology, Division of Rheumatology, Department of Medicine, David Geffen School of Medicine, UCLA; M. Maloney, BA, MA, Medical Student; N. Samed, MD, Resident in Rheumatology; D. Norrie, BSc, MD, Resident, Division of Rheumatology, St. Joseph's Health Care and University of Western Ontario; G. Ouimet, BSc, Research Assistant, Division of Rheumatology, St. Joseph's Health Care; R.D. Hays, PhD, Professor of Medicine, Division of Rheumatology, Department of Medicine, David Geffen School of Medicine and Department of Health Services, School of Public Health, UCLA, and Rand Corporation, Santa Monica, CA.

Address reprint requests to Dr. D. Khanna, Division of Rheumatology, Department of Medicine, David Geffen School of Medicine, 1000 Veterans Avenue, Room 32-59 Rehabilitation Building, Los Angeles, CA 90095. E-mail: dkhanna@mednet.ucla.edu

Accepted for publication July 14, 2008.

Personal non-commercial use only. The Journal of Rheumatology Copyright © 2008. All rights reserved.

approximately 41%.² Fatigue remains an important concern for patients with RA^{3,4} and Outcome Measures in Rheumatoid Arthritis Clinical Trials (OMERACT) recommends inclusion of fatigue as a patient-centered outcome in RA studies⁴.

Recent clinical trials have used fatigue as an outcome measure⁵ and fatigue VAS has been found to be sensitive to change in RA-related clinical outcomes⁶. In addition, Wolfe recently compared the fatigue VAS scale to multi-item fatigue instruments in a large observational study, and showed that VAS scale and multi-item scales had comparable responsiveness statistics and similar correlation coefficients with clinical outcomes in RA⁷. As future clinical trials and observational studies in RA are likely to include a measure of fatigue, it is important to estimate the minimal clinically important difference (MCID) or minimally important difference (MID) — the smallest improvement in score that patients perceive as beneficial and that may lead to a change in the patient's management of fatigue⁸. MID can help clinicians understand whether fatigue score differences between 2 treatment groups are meaningful and if changes within 1 group over time are clinically meaningful⁹. The MID estimates were assessed using an anchor-based approach¹⁰. An “anchor” is a clinically relevant indicator or pointer to which a patient-reported outcome change can be tied^{10,11}. These measures are of clinical relevance and can be “subjective,” such as self-reports of change, or “objective,” such as clinical indicators of response to treatment (joint count or disease severity). Subjective anchors rely on an individual's (subject's or his/her physician's) assessment of the patient's condition. A global rating of change is a well accepted subjective anchor in patient-reported outcome research¹². It is a retrospective assessment of change where a person thinks back to a previous timepoint and states whether he/she has experienced change in a domain of health from that timepoint to present¹³.

We prospectively studied a large clinical practice at a university hospital to determine the MID estimates for fatigue VAS (0–10 scale; for improvement and worsening) using 3 patient-reported anchors — fatigue, pain, and overall health. Previous studies have shown an inherent uncertainty around the MID estimates^{5,14,15} and experts have recommended using several anchors. Our primary anchor was the fatigue scale, but we also performed our analyses using overall and pain anchors, as a previous study has shown an association between pain, overall health, and fatigue.

MATERIALS AND METHODS

Multiple data are collected routinely on patients seen at St. Joseph's Hospital Rheumatology Clinic, which is affiliated with the University of Western Ontario and serves a referral region of approximately 1 million. The data are from patients ($n = 307$) with RA¹⁶ who had at least 2 consecutive visits within 12 months (median duration 5.9 mo) with their rheumatologists. As patients with low disease activity may not have a frequent followup with rheumatologists, we chose the 12-month duration so we can

include all eligible subjects. Data were extracted from medical charts by trained data-extraction persons and entered into a database. Patients completed the fatigue VAS (0–10), which stated, “How much problem has unusual fatigue or tiredness been for you over the past week,” and was anchored from 0 (fatigue is no problem) to 10 (fatigue is a major problem). In addition, patients completed the Health Assessment Questionnaire-Disability Index (HAQ-DI)¹⁷ and a retrospective anchor item at visit 2, “How would you describe your overall fatigue since the last visit?” with response options: Much worsened, Somewhat worsened, Same, Somewhat better, or Much better. Patients who reported Somewhat better or Somewhat worsened at visit 2 were defined as the minimally changed subgroups. The changes in the fatigue VAS scores (time₂ visit – time₁ visit) for the group that reported Somewhat better and Somewhat worsened were determined in order to estimate the MID. This was compared to change scores for the group that reported Same, Much better, and Much worsened. The normality of change in fatigue scores was assessed using the Shapiro-Wilk test. The test was nonsignificant ($p = 0.06$) and we report the MID estimates as mean and 95% confidence intervals (95% CI).

The fatigue change scores were calculated (fatigue from time₂ visit – fatigue from time₁ visit). To assess the usefulness of an anchor, experts recommend reporting correlation between the anchor and changed score; for example, a correlation of zero will make the anchor useless, so a correlation coefficient of 0.30–0.35 has been suggested (for details see Hays, *et al*⁹, Revicki, *et al*¹⁸). We assessed the association between anchor and fatigue change score using the Spearman correlation coefficient (as the anchor is an ordinal variable). The MID was estimated by examining change in the fatigue VAS scores in subjects who were slightly better and slightly worsened. These estimates were compared to those who improved or worsened more than slightly. Responsiveness to change was evaluated using the effect size (ES)¹⁹. ES is ratio of observed change to a measure of variance (also known as signal to noise). For ES, the numerator is the mean change in the fatigue VAS from the baseline to followup and the denominator is the standard deviation of fatigue VAS at baseline ($SD = 2.87$). Cohen's rule of thumb for interpreting ES is that a value of 0.20–0.49 represents a small change, 0.50–0.79 a medium change, and 0.80 or greater a large change¹⁹ and, in general, an ES of 0.2 to 0.5 is usually considered relevant for MID²⁰.

We also estimated MID scores using pain and overall health assessments as anchors. The patients were asked, “How would you describe your pain/overall health since the last visit?” on a scale labeled: Much worsened, Somewhat worsened, Same, Somewhat better, Much better.

In addition, the effect of baseline fatigue VAS score on the MID estimates was estimated. In other words, people with different baseline VAS fatigue scores may require different amounts of improvement or worsening to consider a change to represent an MID¹⁵. We divided the baseline VAS fatigue scores into 2 groups at the median score — the Less severe group (fatigue VAS ≤ 4.5) and the More severe group (fatigue VAS > 4.5).

We also assessed the significance of change at the individual level. The reliable change index (RCI) is a measure to assess the magnitude of change score necessary to be considered statistically reliable and not due to random measurement error, and is a function of the SD and reliability coefficient of an instrument^{21,22}. When RCI is > 1.96 , it is unlikely that the posttest score is due to random measurement error and the change is reliable ($p < 0.05$). RCI is calculated by the following formula:

$$RCI = (\text{individual fatigue score}_{\text{time2}} - \text{individual fatigue score}_{\text{time1}}) / SD_{\text{baseline}} (2 \times [1 - \text{reliability}])^{1/2}$$

We did not assess test-retest reliability of our fatigue VAS. However, a reliability (test-retest) coefficient of 0.70 for the fatigue VAS was found in patients with RA²³. Therefore, we based our primary analysis on the reliability coefficient of 0.70 and performed sensitivity analyses with 0.80 and 0.90. We report the scores that are statistically reliable, and proportion of patients who had individual change greater than the statistically reliable change (improved or worsened) and proportion that had individual changes less than the statistically reliable change.

The data were analyzed using Stata 9.2. $p < 0.05$ was deemed to indicate statistical significance.

RESULTS

The mean (SD) age of the patients was 59.4 (13.2) years, the mean (SD) disease duration was 14.1 (11.5) years, and 256 (83%) were women; the median (25th, 75th percentile) followup between 2 visits was 5.9 (3.9, 7.6) months. Twenty-eight patients (9%) reported no fatigue at baseline and 5 (2%) reported maximum fatigue at baseline (Figure 1). The majority of patients (92%) were using disease-modifying antirheumatic drug (DMARD) therapy, 16% prednisone, 50% nonsteroidal antiinflammatory drugs (NSAID), and 17% were prescribed biologics. The baseline HAQ-DI was 0.84 (SD 0.75), suggesting mild functional disability. The baseline fatigue VAS was 4.2 (SD 2.9) and at followup was 4.3 (SD 2.8) (mean change of -0.07 ± 2.5 ; $p =$ nonsignificant).

The Spearman correlation between the patient assessment of fatigue and change in the fatigue VAS was 0.37 ($p < 0.001$). Of 307 patients, 171 (56%) reported no change at followup visit, 40 (13%) reported being somewhat improved, and 17 (5%) reported being much better. In contrast, 65 (21%) reported being somewhat worsened and 14 (5%) reported being much worsened.

The fatigue change scores for the Somewhat better ($n = 40$) and Somewhat worsened ($n = 65$) groups averaged -1.12 and 1.26 (95% CI -1.86 to -0.37 and 0.68 to 1.86), respectively (Table 1). The change in fatigue scores for those who were much better ($n = 17$) was -1.44 (95% CI $-2.78, -0.10$) and it was $+3.35$ (95% CI $1.94, 4.78$) for much worse ($n = 14$). All the change scores were statistically different ($p < 0.05$) from the score for the no-change group [$n = 171$; score -0.23 (95% CI $-0.57, 0.11$)]. ES estimates were small for the Somewhat better (0.39) and Somewhat worse (0.44) groups, but larger than for the no-change group (ES = 0.08).

For the global and pain anchors, the Spearman correlations between the patient assessment of overall symptoms and pain versus change in the fatigue VAS were both 0.33 ($p < 0.001$). Using the pain anchor, the fatigue change score for Somewhat better ($n = 50$) and Somewhat worsened ($n = 70$) averaged -0.87 and 1.13 (95% CI -1.45 to -0.28 and 0.70 to 1.77), respectively. Using the global anchor, the fatigue change score for Somewhat better ($n = 52$) and Somewhat worsened ($n = 65$) averaged -0.82 and 1.17 (95% CI -1.40 to -0.24 and 0.48 to 1.86), respectively.

Table 2 provides the mean MID scores for improvement in the fatigue scores stratified by the “less severe” (fatigue VAS ≤ 4.5) versus “more severe” (fatigue VAS > 4.5) groups at baseline. The data suggest that patients with “more severe” baseline fatigue scores require a larger change to be minimally improved (MID for the minimally improved group for the low score group = -0.15 , high score group = -3.38). In contrast, the change required for the slightly worsened group with “more severe” fatigue (0.21) was smaller than that for the “less severe” (3.08) fatigue group.

Reliable change index. The amount of change on the fatigue VAS needed to have the RCI to be 1.96 was 3.47 when the reliability coefficient was 0.70. In other words, a change of score from time₁ to time₂ (improvement or worsening) of > 3.47 (on a 0–10 scale) was a real change and was not due to measurement error of the instrument. A change < 3.47 may be real, but a random measurement error cannot be ruled out. Out of 307 patients, 26 (8%) had worsening and 30 (10%) had improvement of scores of > 3.47 . The RCI was 3.14 and 2.64 when the test-retest reliability coefficients were 0.80 and 0.90, respectively.

DISCUSSION

Fatigue is very common in patients with RA^{1,24}, is likely to be multifactorial³, and was recommended as a patient-cen-

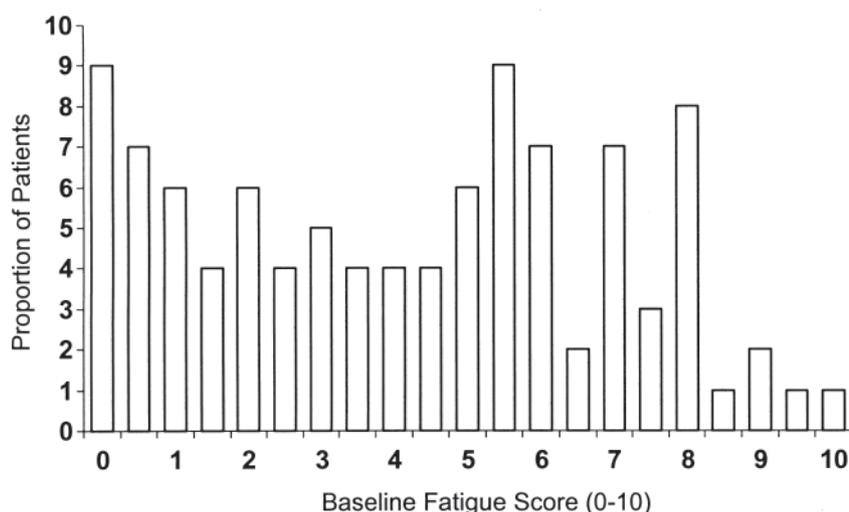


Figure 1. The distribution of baseline fatigue scores on a 0–10 scale.

Table 1. Minimal important difference (MID) estimates for the fatigue VAS (0–10).

Patient-rated Fatigue	Mean Fatigue VAS Change	95% CI	Effect Size	p
Much better (n = 17)	-1.44	-2.78, -0.10	-0.50	0.04
Somewhat better (n = 40)	-1.12	-1.86, -0.37	-0.39	0.03
Same (n = 171)	-0.23	-0.57, 0.11	-0.08	—
Somewhat worse (n = 65)	1.26	0.68, 1.86	0.44	< 0.001
Much worse (n = 14)	3.35	1.94, 4.78	1.17	< 0.001

Negative scores signify improvement and positive scores signify worsening. VAS: visual analog scale. p value denotes comparison of “Same” group to 4 other groups.

Table 2. Mean MID scores for improvement in the fatigue scores stratified by “less severe” fatigue vs “more severe” fatigue at baseline.

Patient-rated Fatigue	n	Fatigue VAS ≤ 4.5 mean score	95% CI	n	Fatigue VAS > 4.5 mean score	95% CI
Much better	11	-0.41	-1.16, 0.34	6	-3.33	-7.05, 0.39
Somewhat better	28	-0.15	-0.86, 0.55	12	-3.38	-4.50, -2.25
Same	102	0.61	0.24, 0.98	69	-1.47	-2.01, -0.94
Somewhat worse	24	3.08	3.08, 0.45	41	0.21	-0.34, 0.76
Much worse	5	5.70	3.66, 7.74	9	2.06	0.69, 3.42

Negative scores signify improvement and positive scores signify worsening. MID: minimal important difference; VAS: visual analog scale.

tered outcome during the OMERACT 8 meeting⁴. Work by Kirwan and Hewlett¹ and Wolfe and Michaud³ has shown that fatigue is an important concern for patients with RA.

We found that our baseline mean fatigue score of 4.2 (SD 2.9) is very similar to a mean score of 4.2 (SD 2.8) seen by Wolfe in 7760 patients with RA⁷. Our data add to the literature by comparing various patient-reported outcome anchors and assessing the MID overall in fatigue and in high and low fatigue. Our MID estimates for improvement ranged from -0.82 to -1.12 for improvement and 1.13 to 1.26 for worsening. Previous studies have shown an inherent uncertainty around the MID estimates^{5,14,15} and experts have recommended using several anchors. Our primary anchor was the fatigue scale, but we also performed our analyses using overall and pain anchors, as previous studies have shown an association between pain, overall health, and fatigue. Our results were very similar using the 3 anchors.

Wells, *et al* assessed MID estimates for the fatigue VAS scale in 2 randomized controlled studies comparing abatacept versus placebo in RA and found similar results; their MID estimates for improvement ranged from 6.7 to 17.0 for 2 studies on a 0–100 scale (or 0.67 to 1.70 on a 0–10 scale)⁵. Apart from differences in our study design and that of Wells, *et al*, we used external anchors to determine our MID estimates, whereas Wells, *et al* used internal anchors (change in HAQ-DI, pain VAS, and patient global VAS) based on previous studies. They followed their estimates with a Delphi exercise and reached a consensus estimate of 10 (on a 0–100 scale).

Previous literature has shown that an effect size of 0.20

to 0.50 corresponds to MID for a patient-reported outcome measure^{14,20}. The ES of the Somewhat better and Somewhat worse groups were small (0.39 and 0.44, respectively), supporting the MID results²⁰. In addition, the MID estimates were in the right direction and of larger magnitude than that for the no-change group.

As noted^{10,15}, the MID estimates may depend on the baseline scores. This trend was seen in our analysis (Table 2), where people with higher baseline scores required a larger change in their fatigue VAS for improvement to be considered as minimally improved. Conversely, patients with lower baseline scores required a larger change in their fatigue VAS for worsening. This may be related to floor and ceiling effect (where people near the bottom of the scale are limited by how much they can improve or near the top of the scale are limited by how much they can worsen) or may represent difference in interpretation of the scale along the continuum²⁵.

We introduce the concept of statistically significant change in individual patient-reported outcome scores rather than group scores. RCI indicates whether change for an individual patient is beyond the measurement error of the instrument; RCI has the advantage of yielding a direct test of the significance of individual change²¹. As expected, clinically significant change was larger for individual patients (3.47) compared to group change (-0.82 to -1.12 for improvement and 1.13 to 1.26 for worsening) and RCI decreased as the test-retest reliability of the VAS increased.

Our study is not without limitations. First, the global ratings of change ask people to remember how their health was

at last visit and the retrospective self-reports are known to be subject to recall bias¹³. Researchers have proposed prospective anchors and they should be studied in the future¹³. When using a prospective anchor, a person rates his/her pain, for example, at time₁ and time₂: mild, moderate, moderately severe, severe, very severe, or unbearable. Those reporting a change over the 2 timepoints (e.g., “moderate” at time₁ and “mild” at time₂) constitute the minimal change subgroup. Second, the MID estimates are based on patients followed at one center and may need to be assessed in different clinical settings before the estimates can be generalized. Our patients were older than the average RA disease onset age of 50 years and had mild functional disability as assessed by the HAQ-DI²⁶. The MID estimates may differ in younger patients and in those with moderate to high disability.

These limitations notwithstanding, we provide MID estimates for fatigue for improvement and worsening that can aid in interpreting fatigue VAS in day-to-day care in clinical practice and which may be relevant for sample size calculations in observational studies.

REFERENCES

- Kirwan JR, Hewlett S. Patient perspective: reasons and methods for measuring fatigue in rheumatoid arthritis. *J Rheumatol* 2007;34:1171-3.
- Wolfe F, Hawley DJ, Wilson K. The prevalence and meaning of fatigue in rheumatic disease. *J Rheumatol* 1996;23:1407-17.
- Wolfe F, Michaud K. Fatigue, rheumatoid arthritis, and anti-tumor necrosis factor therapy: an investigation in 24,831 patients. *J Rheumatol* 2004;31:2115-20.
- Kirwan JR, Minnock P, Adebajo A, et al. Patient perspective: fatigue as a recommended patient centered outcome measure in rheumatoid arthritis. *J Rheumatol* 2007;34:1174-7.
- Wells G, Li T, Maxwell L, Maclean R, Tugwell P. Determining the minimal clinically important differences in activity, fatigue, and sleep quality in patients with rheumatoid arthritis. *J Rheumatol* 2007;34:280-9.
- Wells GA, Li T, Maxwell L, Maclean R, Tugwell P. Responsiveness of patient reported outcomes including fatigue, sleep quality and activity limitation and quality of life following treatment with abatacept in patients with rheumatoid arthritis. *Ann Rheum Dis* 2008;67:260-5.
- Wolfe F. Fatigue assessments in rheumatoid arthritis: comparative performance of visual analog scales and longer fatigue questionnaires in 7760 patients. *J Rheumatol* 2004;31:1896-902.
- Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertain the minimal clinically important difference. *Control Clin Trials* 1989;10:407-15.
- Hays RD, Farivar S, Liu H. Approaches and recommendations for estimating minimally important differences for health-related quality of life measures. *COPD* 2005;2:63-7.
- Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. *J Clin Epidemiol* 2003;56:395-407.
- Khanna D, Tsevat J. Health-related quality of life — an introduction. *Am J Manag Care* 2007;13 Suppl:S218-23.
- Guyatt GH, Norman GR, Juniper EF, Griffith LE. A critical look at transition ratings. *J Clin Epidemiol* 2002;55:900-8.
- Yost KJ, Sorensen MV, Hahn EA, Glendenning GA, Gnanasakthy A, Cella D. Using multiple anchor- and distribution-based estimates to evaluate clinically meaningful change on the Functional Assessment of Cancer Therapy-Biologic Response Modifiers (FACT-BRM) instrument. *Value Health* 2005;8:117-27.
- Khanna D, Furst DE, Wong WK, et al. Reliability, validity, and minimally important differences of the SF-6D in systemic sclerosis. *Qual Life Res* 2007;16:1083-92.
- Khanna D, Furst DE, Hays RD, et al. Minimally important difference in diffuse systemic sclerosis — Results from the D-Penicillamine Study. *Ann Rheum Dis* 2006;65:1325-9.
- 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.
- Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. *Arthritis Rheum* 1980;23:137-45.
- Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol* 2008;61:102-9.
- Cohen J. A power primer. *Psychol Bull* 1992;112:155-9.
- Sloan JA, Cella D, Hays RD. Clinical significance of patient-reported questionnaire data: another step toward consensus. *J Clin Epidemiol* 2005;58:1217-9.
- Ferguson RJ, Robinson AB, Splaine M. Use of the reliable change index to evaluate clinical significance in SF-36 outcomes. *Qual Life Res* 2002;11:509-16.
- Hays RD, Brodsky M, Johnston MF, Spritzer KL, Hui KK. Evaluating the statistical significance of health-related quality-of-life change in individual patients. *Eval Health Prof* 2005;28:160-71.
- Kvien TK, Mowinckel P, Heiberg T, et al. Performance of health status measures with a pen based personal digital assistant. *Ann Rheum Dis* 2005;64:1480-4.
- Hewlett S, Hehir M, Kirwan JR. Measuring fatigue in rheumatoid arthritis: a systematic review of scales in use. *Arthritis Rheum* 2007;57:429-39.
- Hays RD, Woolley JM. The concept of clinically meaningful difference in health-related quality-of-life research. How meaningful is it? *Pharmacoeconomics* 2000;18:419-23.
- Krishnan E, Sokka T, Hakkinen A, Hubert H, Hannonen P. Normative values for the Health Assessment Questionnaire disability index: benchmarking disability in the general population. *Arthritis Rheum* 2004;50:953-60.