

The Minimally Important Difference for the Health Assessment Questionnaire in Rheumatoid Arthritis Clinical Practice Is Smaller Than in Randomized Controlled Trials

JANET E. POPE, DINESH KHANNA, DEBORAH NORRIE, and JANINE M. OUIMET

ABSTRACT. *Objective.* Patient-reported outcomes are used in clinical practice and trials. We studied a large clinical practice to determine the minimally important difference (MID) estimates for (1) the Health Assessment Questionnaire–Damage Index (HAQ-DI): improvement and worsening using patient global assessment anchor; and (2) pain using a patient-reported pain anchor.

Methods. Patients with rheumatoid arthritis (RA; N = 225) had clinic visits at 2 timepoints within 1 year, completed the HAQ-DI and pain visual analog scale (VAS; 0–100 mm), and answered the question, “How would you describe your overall status/overall pain since the last visit?”, as much worsened, somewhat worsened, the same, somewhat improved, or much improved. If rated as somewhat improved or worsened, they were defined as the minimally changed subgroups.

Results. Eighty-three percent were women, mean age 60 years, with disease duration 11.7 ± 10.7 years. The baseline HAQ-DI was $0.97 \pm \text{SD } 0.76$, and at followup 1.0 ± 0.77 (mean change $+0.03 \pm 0.40$). The baseline pain VAS was 42.3 ± 28.8 , and at followup 38.5 ± 27.9 (mean change -2.8 ± 25.9). The mean (SD) HAQ-DI change score was -0.09 (0.42) for somewhat improved and 0.15 (0.33) for somewhat worsened. The HAQ-DI change for somewhat/much better was -0.20 ± 0.52 , and for somewhat/much worse $+0.21 \pm 0.33$. For pain, somewhat improved changed by -11.9 mm on the VAS, and somewhat worsened by 6.8 mm. Estimates for HAQ-DI and pain were larger than the for no-change group, 0.03 (0.32) and -3.2 (20.9).

Conclusion. The MID for HAQ-DI in clinical practice is smaller than it is in trials. This may have implications for observational studies and clinical care. (First Release Dec 15 2008; J Rheumatol 2009;36:254–9; doi:10.3899/jrheum.080479)

Key Indexing Terms:

MINIMALLY IMPORTANT DIFFERENCE
HEALTH ASSESSMENT QUESTIONNAIRE-DISABILITY INDEX
RHEUMATOID ARTHRITIS
RANDOMIZED CONTROLLED TRIALS

The Stanford University Health Assessment Questionnaire-Disability Index (HAQ-DI) is a musculoskeletal-targeted self-report tool that assesses the functional status for per-

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

D. Norrie was supported by an unrestricted research grant from the Department of Medicine, the University of Western Ontario, London, Canada.

J.E. Pope, MD, MPH, FRCPC, Division of Rheumatology, Department of Medicine, University of Western Ontario; D. Khanna, MD, MSc, Division of Rheumatology, Department of Medicine, David Geffen School of Medicine, and Department of Health Services, School of Public Health, University of California at Los Angeles; D. Norrie; J. Ouimet, MSc, Division of Rheumatology, Department of Medicine, University of Western Ontario.

Address reprint requests to Dr. J. Pope, St. Joseph's Health Care, 268 Grosvenor Street, PO Box 5777, London, Ontario N6A 4V2.

Accepted for publication September 18, 2008.

forming activities of daily living; it is scored from 0 (no disability) to 3 (severe disability), representing the averaging of the worst score in each of 8 domains of daily function¹. It is the most widely used functional outcome measurement in randomized controlled trials of rheumatoid arthritis (RA). The HAQ-DI is feasible, reliable, and valid for clinical trials in RA and is sensitive to clinically relevant changes in function². Higher scores (> 1.0) predict worse outcomes³.

Minimally important difference (MID) is the smallest difference in a measure that patients perceive as a change⁴. It is the smallest change in a score that would be clinically relevant, and MID are useful for determining sample sizes for future studies. In order to determine the MID, an external anchor such as a physician or patient global assessment is used. The MID may reflect either an improvement or a worsening. The minimal clinically important improvement is defined as the smallest change in measurement that signifies an important improvement. A related concept is the Patient Acceptable Symptom State (PASS), which has been defined as the highest level of symptom beyond which

patients consider themselves well⁵. Changes of 10% or roughly 10 on a 100 mm visual analog scale (VAS) correspond to MID for patient-reported measures across many studies⁶.

In cross-sectional studies of patients with RA, the MID for HAQ-DI improvement was -0.19 to -0.23 ^{7,8}, whereas it was $+0.49$ ⁷ for worsening. In RA clinical trials the MID in HAQ-DI improvement ranged from -0.22 to -0.24 ^{7,9,10}. In trials, patients perceive worsening sooner than they do improvement, causing the MID for deterioration to be smaller than those for improvement¹¹. This could also occur in clinic patients hoping to “not deteriorate,” or in other words an expectation bias may be present.

We prospectively studied a large clinical practice at a university hospital to determine (1) the MID estimates for HAQ-DI for improvement and worsening (i.e., the minimally clinically important improvement and the minimally clinically important deterioration) using a patient global assessment anchor; and (2) the MID estimates for pain (for improvement and worsening) using a patient-reported pain anchor. We hypothesized (1) that patients with RA in a clinical practice would have lower MID scores than those seen in randomized controlled trials (RCT); (2) that MID scores would be different bidirectionally (improvement and worsening); and (3) that MID scores would vary according to baseline HAQ and possibly disease duration. We speculated that RA patients with higher HAQ-DI disability could require greater change to constitute MID. We did not include the PASS in this study as we did not ask patients if they were in an acceptable state at the time of their visits.

MATERIALS AND METHODS

Multiple data are collected routinely on patients seen at St. Joseph's Hospital Rheumatology Clinic, affiliated with the University of Western Ontario, which services a referral region of roughly 1 million people. The data are from patients seen by one rheumatologist (JP) serially who had been diagnosed with RA meeting the American College of Rheumatology criteria¹² seen over a 6-month period ($n = 347$) and who had at least 2 consecutive visits within 12 months. One hundred twenty-two patients did not have complete data and were excluded from the analysis, leaving a final sample of 225 patients. Data were extracted from medical charts by a trained data-extraction person (DN) and entered into a database. Patients completed the HAQ-DI, pain VAS (0–100 mm), anchored from 0 (no pain at all) to 100 (very severe pain), and completed a 5-point Likert scale for the question, “How would you describe your overall status since the last visit?” on a scale labeled much worsened, somewhat worsened, the same, somewhat improved, much improved. Similarly, the patients were asked to complete a 5-point Likert scale for the question, “How would you describe your overall pain since the last visit?” on a scale labeled much worsened, somewhat worsened, the same, somewhat improved, much improved. Patients who reported somewhat improved or somewhat worsened were defined as the minimally changed subgroups. The changes in the HAQ-DI and pain VAS scores for the group that slightly improved and slightly worsened at their next followup visit within 1 year were determined in order to estimate the MID. This was compared to change scores for the group that reported the same, much improved, and much worsened.

The HAQ-DI change scores were calculated (HAQ from most recent visit minus HAQ from previous visit). We also collapsed categories in an

exploratory analysis where HAQ-DI change scores corresponding to becoming better/much better or worse/much worse were compared to reporting no change in overall disease status.

To assess the usefulness of an anchor, change in the anchor and change in the health-related quality of life (HRQOL) scores should have a correlation coefficient of at least > 0.37 ¹³. The correlation coefficient of > 0.37 corresponds to an effect size of 0.80 (a large effect, as proposed by Cohen¹⁴). This was assessed using the Spearman correlation coefficient (as change in anchor is an ordinal variable) between the patient global and pain assessments (the anchors) and change in the HAQ-DI and pain scores. The MID was estimated by examining change in the HAQ-DI and pain VAS scores in subjects who were slightly improved 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). ES is the mean change in the HAQ-DI or pain VAS from the baseline to followup divided by the standard deviation at baseline (0.76 for HAQ-DI and 28.8 for pain VAS). Cohen's rule of thumb for interpreting effect size is that a value of 0.20–0.49 represents a small change, 0.50–0.79 a medium change, and ≥ 0.80 a large change, and in general an ES of 0.2 to 0.5 is usually considered relevant for MID¹⁴.

As an exploratory analysis, the effect of baseline HAQ-DI score and pain VAS score on the MID estimates was assessed. In other words, people with different baseline HAQ-DI or VAS pain scores may require different amounts of improvement or worsening to consider a change as representing an MID. We divided the baseline HAQ-DI and VAS pain scores into 2 groups — mild group (HAQ-DI < 1.0 and pain VAS ≤ 33.3) and the moderate to severe group (HAQ-DI ≥ 1.0 and pain VAS > 33.4).

RESULTS

The 225 patients with RA had long disease duration (11.7 yrs) and an average age of 60 years and 83% were women. The majority of patients (92%) were using disease-modifying antirheumatic drug (DMARD) therapy, 16% prednisone, 50% nonsteroidal antiinflammatory drugs (NSAID), and 17% were prescribed biologic therapies. The average baseline HAQ-DI was 0.97 ± 0.76 and at followup it was 1.0 ± 0.77 , with a mean change of $+0.03 \pm 0.40$ (Table 1).

Of the 225 patients, 53% reported no change at followup visit, 16% reported being somewhat improved, and 5% reported being much improved. In contrast, 22% reported being somewhat worsened and 4% reported being much worsened on the global assessment anchor. The Spearman correlation coefficient between the patient assessment of global change and the change in the HAQ-DI was 0.36 ($p < 0.001$).

Table 1. Baseline characteristics of the study population (N = 225).

Characteristic	Min, Maximum	
Age, mean (SD) yrs	60.5 (13.6)	17, 90
Female, %	83	—
Years with RA, mean (SD)	11.7 (10.7)	0, 57
% using DMARD	92	—
% using steroids	16	—
% using NSAID	50	—
% using biologics	17	—
HAQ-DI (0–3), mean (SD)	0.97 (0.76)	0, 2.75
Pain VAS score (0–100), mean (SD)	42.3 (28.8)	0, 100

SD: standard deviation, DMARD: disease modifying antirheumatic drugs, NSAID: nonsteroidal antiinflammatory drugs, HAQ-DI: Health Assessment Questionnaire-Disability Index, VAS: visual analog scale.

The HAQ-DI change score for the group that somewhat improved and somewhat worsened was mean -0.09 (SD 0.42) and 0.15 (SD 0.33), respectively (Table 2). This corresponds to an ES of 0.12 and 0.20. This difference was also statistically significant compared to the no-change group [HAQ-DI mean 0.03 (SD 0.32), ES 0.04; Table 2]. The HAQ-DI change for somewhat better/much better was -0.20 ± 0.52 (ES 0.27) and $+0.21 \pm 0.33$ (ES 0.27) for somewhat worse/much worse, respectively, whereas the mean change in HAQ-DI for “the same” was $+0.03 \pm 0.32$.

Pain VAS results. Two hundred twelve patients completed the pain VAS at 2 timepoints. The baseline average pain VAS was 42.3 ± 28.8 and at followup 38.5 ± 27.9 , with a mean change of -2.8 ± 25.9 (Table 3). Of the 212 patients, 52% reported no change, 14% reported being somewhat improved, and 5% reported being much improved. In contrast, 24% reported being somewhat worse and 5% reported being much worse. This distribution was similar to the HAQ-DI. The Spearman correlation coefficient between the patient assessment of pain change and change in the pain VAS was 0.37 ($p < 0.001$).

On average, patients who answered “somewhat better” and “somewhat worse” on the pain anchor changed by -11.9 and 6.8 mm on the VAS, corresponding to an ES of 0.41 and 0.23, respectively (Table 3). These estimates were larger than for the no-change group [-3.2 (SD 20.9); ES 0.11].

Table 4 provides the mean MID scores for improvement in the HAQ-DI scores stratified by severity of the mild

(HAQ-DI < 1.0) compared to the moderate/severe (HAQ-DI ≥ 1.0) groups at baseline. Although they are exploratory, the data suggest that patients with moderate to severe baseline HAQ-DI scores require a larger improvement to be minimally improved, as exemplified by the HAQ-DI (MID for the minimally improved group for the low-score group = -0.03 , and for the high-score group = -0.14). In contrast, the change required for the slightly worsened group with moderate to severe HAQ-DI (0.09) was smaller than that for the mild HAQ-DI group (0.24). A similar pattern was seen with the pain VAS. We did not have enough early RA disease duration data to compare HAQ-DI changes between early and late disease.

We also investigated the proportion of changes in patient-reported global states that were concordant with change in HAQ (Table 5). As might be expected, for larger changes such as much better (73% concordance) or much worse (100% concordance) there was more agreement. However, in better, the same, and worse, the HAQ was not as often concordant, as the usual HAQ-DI changes were small.

DISCUSSION

MID estimates provide a benchmark for future design of RA clinical and observational trials by helping researchers and clinicians understand whether HRQOL score differences between 2 treatment groups are meaningful, or if changes within one group over time are meaningful¹⁵. On average, our MID scores for improvement and worsening of HAQ-DI

Table 2. Minimal important difference (MID) estimates for the HAQ-DI.

Patient-rated Overall Status	HAQ-DI Change, mean (SD)	95% CI	Effect Size	p
Much better (n = 11)	-0.57 (0.67)	-1.01 to -0.12	0.74	< 0.0001
Somewhat better (n = 35)	-0.09 (0.42)	-0.23 to 0.05	0.12	0.08
Same (n = 120)	0.03 (0.32)	-0.030 to 0.09	0.04	—
Somewhat worse (n = 50)	0.15 (0.33)	0.060 to 0.25	0.20	0.02
Much worse (n = 9)	0.50 (0.13)	0.40 to 0.60	0.65	< 0.0001

HAQ-DI scored from 0 (no disability) to 3 (severe). HAQ-DI change score = HAQ score at most recent visit – score at previous visit. Negative score indicates improvement.

Table 3. Minimal important difference estimates for the pain visual analog scale (VAS; 0–100 mm).

Patient-rated Overall Status	Pain VAS Change, mean (SD)	95% CI	Effect Size	p
Much better (n = 10)	-42.1 (34.2)	-66.5 to -17.6	1.46	< 0.001
Somewhat better (n = 30)	-11.9 (30.8)	-23.4 to -0.4	0.41	0.07
Same (n = 110)	-3.2 (20.9)	-7.1 to 0.8	0.11	—
Somewhat worse (n = 51)	6.8 (22.7)	0.4 to 13.1	0.23	0.007
Much worse (n = 11)	15.3 (19.5)	2.2 to 28.4	0.53	0.005

0 is no pain, 100 is “very severe pain”. Pain VAS change score = Pain VAS score at most recent visit – score at previous visit. Negative score indicates improvement.

Table 4. Mean (SD) MID scores for improvement in the HAQ-DI and pain VAS.

HAQ-DI Baseline Score					
Mean MID (SD) in the slightly improved group (n = 17)	HAQ-DI < 1.0	Mean MID (SD) in the slightly worse group (n = 21)	Mean MID (SD) in the slightly improved group (n = 18)	HAQ-DI ≥ 1.0	Mean MID (SD) in the slightly worse group (n = 29)
	Mean (SD) in no-change group (n = 77)			Mean (SD) in no-change group (n = 43)	
-0.03 (0.19)	0.08 (0.27)	0.24 (0.31)	-0.14 (0.56)	-0.07 (0.38)	0.09 (0.37)

Pain VAS Baseline Score					
Mean MID (SD) in the slightly improved group (n = 10)	Pain VAS ≤ 33.3	Mean MID (SD) in the slightly worse group (n = 16)	Mean MID (SD) in the slightly improved group (n = 20)	Pain VAS > 33.4	Mean MID (SD) in the slightly worse group (n = 35)
	Mean (SD) in no-change group (n = 61)			Mean (SD) in no-change group (n = 49)	
-12.6 (26.3)	7.3 (17.6)	17.2 (20.8)	-24.2 (25.5)	-16.3 (16.9)	2.0 (22.2)

Table 5. Proportion of change scores (%) concordant with directional trend of category.

Overall Status Category (direction)	N	Negative HAQ Change (Improved HAQ) (-)	Neutral HAQ Change (No change in HAQ) (0)	Positive HAQ Change (Worse HAQ) (+)
Much better (-)	11	73	9	18
Better (-)	35	40	40	20
Same (0)	120	34	24	42
Worse (+)	50	18	18	64
Much Worse (+)	9	0	0	100

* Negative HAQ change score indicates improvement; positive score indicates worsening.

were -0.09 and 0.15, respectively, scores that are lower than those seen previously^{7,9,10}. In RA studies the MID in HAQ-DI improvement has ranged from -0.22 to -0.24^{7,9,10}. Using the HAQ-DI > 1.0 at baseline, a representation of patients with moderate to severe disease in RA studies, the MID was 0.14 for improvement and 0.09 for worsening (as compared to 0.22 for improvement). After we combined the groups that improved “somewhat” and “a lot,” the average change in HAQ-DI was -0.20, a score similar to the MID HAQ-DI score used in other RA studies.

The difference in MID in clinical practice and RCT in RA has been alluded to by others¹⁶ and could be due to several reasons. First, the majority of patients in clinical trials have moderate to severe disease. Clinical trials usually recruit patients with active disease, and the patients are started on either active treatment or placebo to achieve better control of the disease. The improvement may be influenced in part by a statistical phenomenon, the regression to the mean, thus causing the need for a larger change in HAQ actual change (and placebo effect). In contrast, our group of patients was followed in a rheumatology practice, where DMARD and other agents were added as deemed necessary;

the majority of patients were stable taking their current DMARD. Second, the expectations of patients in RCT, in addition to taking active versus control treatment, are different than in routine practice, which could give an added placebo effect. Third, the HAQ-DI is explained by both reversible and irreversible components (disease activity and damage) as recently demonstrated¹⁷. Thus, disease duration can affect the HAQ-DI, where early in the disease course the reversibility is greater, as there is little or no accrued damage, and RA is primarily manifested as inflammation with swollen joints; and later, radiographic damage and deformities explain more of the HAQ-DI. Therefore, some residual HAQ cannot be altered (due to damage). The long mean disease duration in our patients may explain the slight overall mean worsening of the HAQ-DI in the interval that was studied. However, MID for worsening were still less than those for improvement in the groups with baseline HAQ > 1.0 and pain VAS > 33.4; thus, these patients with greater disease activity had results more consistent with populations enrolled in trials¹¹.

Our data on pain using a 100 mm VAS demonstrate that scores of -11.9 and 6.8 are indicative of MID for improve-

ment and worsening using the pain anchor. This result is consistent with data from Wells, *et al*⁷, where a change of 10 mm on a 0–100 mm VAS was suggestive of the MID score. Wyrwich, *et al* have shown that MID anchored to patient global assessments changed by 0.5 SD¹⁸⁻²⁰. In our study population identification of MID by the 0.5-SD rule works, with the exception of the “slightly improved” group in HAQ scores in both categories. We chose overall status as the anchor as it seemed most closely related to RA self-assessed disease. One would think the lack of specificity of the anchor (i.e., not an anchor of overall disability or function) would increase the MID due to other “noise” in self-rated overall status.

As previously noted¹³, the MID estimates may depend on the baseline scores. This trend was seen in our analysis (Table 4), where people with higher baseline scores (defined as baseline score of HAQ-DI ≥ 1.0 and pain VAS > 33.4) required a larger improvement in their HAQ-DI and pain VAS scores to be considered as minimally improved by the patients. Conversely, people with lower baseline scores (defined as baseline score of HAQ-DI < 1.0 and pain VAS ≤ 33.4) required a greater worsening in their HAQ-DI and pain VAS scores to be considered as minimally worsened by the patients. We are uncertain why this is the case. One would predict larger changes in both improved and worsened HAQ-DI and VAS pain at higher baseline values (and thus percentage changes would be similar). In general, at lower scores, RA patients may be more optimistic and require larger changes for worsening than improving.

Our study has some limitations. First, the patients were primarily Caucasian and had to be able to read English in order to complete their forms. Second, most patients had long disease duration (mean 11.7 yrs), so the results may not be generalizable to very early RA, as the HAQ-DI is related to both inflammation and damage. Third, the time reference for the HAQ-DI is “over the last week” and we asked for the Likert scale “since the last visit,” which was usually 6 to 8 months previously; the retrospective ratings are susceptible to recall biases¹⁵. In addition to forgetting, recall has been shown to be influenced by more salient events and by one’s current mood state¹⁵, and the long duration between visits may have been a limitation. Although patients often ignore time anchors on questions, delay could have increased recall bias, which could have resulted in increased random error and, therefore, increased or decreased MID. Thus, the MID is susceptible to recall bias, as are all repeated questionnaires in rheumatology. Fourth, the choice of anchors can influence the MID scores. We used anchors that had a moderate correlation with HRQOL measures, providing some confidence in our results. Fifth, although the patients were from the practice of a single rheumatologist, treatment for this study was not standardized, and future studies should confirm these results. We could also include the Patient Acceptable Symptom State in future studies. Finally, elimi-

nating some patients with incomplete data in practice may have affected the results. For instance, the patients who did not complete the HAQ may not have been fully literate in English, which could bias the results.

One could question the methods we used to calculate MID, or if, looking at mean change of HAQ in each global health state (the same, better, worse, much better, and much worse), there should be a correction for the amount of change in HAQ that the “no change” group has. For instance, the HAQ-DI no-change group has a mean change of 0.03, so our better change of 0.09 could have a correction of 0.03 (absolute value) added, which would be 0.12 (still smaller than what has been reported in RCT), and similarly for worsening, 0.03 added to 0.15 would be 0.18. In the MID literature, it is important that an anchor is used that is relevant and that is correlated to the variable of interest, so our methodology, although novel, meets the MID criteria. We are investigating the MID in clinical practice further by asking patients if they think they are the same in the variable of interest or have changed a little or a lot.

Often over time HAQ scores worsen, but patients may say they are stable or even improved; however, over the time between one visit and the next, there can be changes in disease state that are important and may be better, or worse, despite many patients being stable. Also, almost half of the patients who change by a HAQ score of 0.25 will rate themselves as unchanged. It is important to note that our observations are averages and that individual patients have changes in HAQ that are incremental (e.g., 0, 0.125, 0.25, 0.375, etc.), so the MID in individuals are different than group averages.

In summary, the MID scores for HAQ-DI in a clinical practice were smaller than those seen in previous clinical trials. The MID scores for pain VAS were similar to those observed in another study. The MID scores are influenced by baseline HRQOL scores, and may be influenced by disease duration, and our results from this study do not comment on very early RA. In addition, the MID changes are different for worsening (usually needing a larger value) than for improving. We observed that the MID for deterioration were much less than for improvement in patients with more pain and impairment in physical function.

REFERENCES

1. Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. *Arthritis Rheum* 1980;23:137-45.
2. 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.
3. Wolfe F, Michaud K, Gefeller O, Choi HK. Predicting mortality in patients with rheumatoid arthritis. *Arthritis Rheum* 2003;48:1530-42.

4. Sloane JA. Assessing the minimally clinically significant difference: scientific considerations, challenges and solutions. *COPD* 2005;2:57-62.
5. Kvien TK, Heiberg T, Hagen KB. Minimal clinically important improvement/difference (MCII/MCID) and patient acceptable symptom state (PASS): what do these concepts mean? *Ann Rheum Dis* 2007;66 Suppl:iii40-iii41.
6. Norman GR, Sloane 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.
7. Wells GA, Tugwell P, Kraag GR, Baker PR, Groh J, Redelmeier DA. Minimum important difference between patients with rheumatoid arthritis: the patient's perspective. *J Rheumatol* 1993;20:557-60.
8. Redelmeier DA, Lorig K. Assessing the clinical importance of symptomatic improvements. An illustration in rheumatology. *Arch Intern Med* 1993;153:1337-42.
9. Kosinski M, Zhao SZ, Dedhiya S, Osterhaus JT, Ware JE. Determining minimally important changes in generic and disease-specific health-related quality of life questionnaires in clinical trials of rheumatoid arthritis. *Arthritis Rheum* 2000;43:1478-87.
10. 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.
11. Strand V, Crawford B. Improvement in health-related quality of life in patients with SLE following sustained reductions in anti-dsDNA antibodies. *Exp Rev Pharmacoeconomics Outcomes Res* 2005;5:317-26.
12. 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.
13. 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. Epub 2006 Mar 15.
14. Cohen J. A power primer. *Psychol Bull* 1992;112:155-9.
15. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. *J Clin Epidemiol* 2003;56:395-407.
16. Wolfe F, Michaud K, Strand V. Expanding the definition of clinical differences: from minimally clinically important differences to really important differences. Analyses in 8931 patients with rheumatoid arthritis. *J Rheumatol* 2005;32:583-9.
17. Aletaha D, Smolen J, Ward MM. Measuring function in rheumatoid arthritis: Identifying reversible and irreversible components. *Arthritis Rheum* 2006;54:2784-92.
18. Wyrwich KW. Minimal important difference thresholds and the standard error of measurement: is there a connection? *J Biopharm Stat* 2004;14:97-110.
19. 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.
20. Wyrwich KW, Nienaber NA, Tierney WM, Wolinsky FD. Linking clinical relevance and statistical significance in evaluating intra-individual changes in health-related quality of life. *Med Care* 1999;37:469-78.