Determining the Minimal Clinically Important Differences in Activity, Fatigue, and Sleep Quality in Patients with Rheumatoid Arthritis

GEORGE WELLS, TRACY LI, LARA MAXWELL, ROSS MacLEAN, and PETER TUGWELL

ABSTRACT. ***Objective.*** To determine the minimal clinically important differences (MCID) in the patient-reported outcomes of activity (0–30, number of days of limitation), fatigue (0 = none, 100 = complete), and sleep quality (0 = no problems, 100 = worst case) for patients with rheumatoid arthritis (RA).

Methods. Two randomized controlled trials comparing abatacept to placebo in RA patients were considered: ATTAIN (n = 391) and AIM (n = 652). An internal anchor-based approach was used to derive the MCID using the Health Assessment Questionnaire, patient global assessment, and pain as anchors. Minimal important change in activity, fatigue, and sleep were determined by estimating mean changes in these outcomes in patients showing change in a narrow range about the MCID of the internal anchor. Correlation analysis was used to determine the consistency of the changes in the outcomes and anchors, and a Delphi process was used to determine the final MCID values.

Results. For the 2 trials, consistent patterns of change for activity, fatigue, and sleep and the internal anchors were found with correlations in the range of 0.5, 0.7, and 0.4, respectively. The mean changes for activity, fatigue, and sleep in a narrow range about the MCID of the 3 internal anchors corresponding to the 2 trials were: 3.4 to 4.3 for activity; 6.7 to 17.0 for fatigue; and 4.1 to 7.3 for sleep. Following the Delphi process the MCID determined were 4 for activity, 10 for fatigue, and 6 for sleep.

Conclusion. These MCID for activity limitation, fatigue, and sleep problems can be used in designing clinical trials and providing benchmarks in assessing patient improvement. (J Rheumatol 2007;34:280–9)

Key Indexing Terms:
MINIMAL CLINICAL IMPORTANT DIFFERENCE RHEUMATOID ARTHRITIS
ACTIVITY FATIGUE SLEEP QUALITY

Assessing patient-reported outcomes (PRO) is important in clinical trials where 2 treatments may have similar effects in controlling or curing disease but different effects on symptoms, function, or other quality of life issues. At the OMERACT 6 Low Disease Activity State (LDAS) Workshop for rheumatoid arthritis (RA), patients at the conference critically reviewed and discussed this concept with the goal to ensure that any definition of LDAS will take into consideration the patient perspective and be ultimately acceptable to patients[1]. The 2 outcomes that patients indicated were of significant importance were fatigue and sleep. These 2 patient-reported outcomes, as well as activity limitation, are the focus of this report.

Challenges arise when determining the clinical significance of any change or difference observed in an outcome measure and in developing a single definition of response indicating a patient has or has not improved. The focus is often on the determination of minimal clinically important differences (MCID). MCID can be considered as the smallest change or difference in an outcome measure that is perceived as beneficial and that would lead to a change in the patient's medical management, assuming an absence of excessive side effects and costs. In determining MCID for an outcome measure, several ingredients are needed: an indicator that change has occurred/a difference exists; an important observed

From the Department of Epidemiology and Community Medicine, University of Ottawa; Bristol-Myers Squibb, Princeton, New Jersey, USA: and the Institute of Population Health, University of Ottawa, Ottawa, Ontario, Canada.

Supported in part by an unrestricted research grant-in-aid from Bristol-Myers Squibb.

G. Wells, PhD, Department of Epidemiology and Community Medicine, University of Ottawa; T. Li, PhD, Bristol-Myers Squibb; L. Maxwell, BSc, Institute of Population Health, University of Ottawa; R. MacLean, MD, Bristol-Myers Squibb; P. Tugwell, MD, MSc, Institute of Population Health, University of Ottawa.

Address reprint requests to Prof. G. Wells, Department of Epidemiology and Community Medicine, University of Ottawa, 451 Smyth Road, Ottawa, Ontario KIH 8M5, Canada.

Accepted for publication October 15, 2006.

change or difference based on a valid assignment of importance; and an appropriate method to determine the threshold level within the distribution of important change or difference scores.

We conducted an extensive search to retrieve all relevant articles related to specific topics on MCID, as well as to identify any methodology articles in the medical literature[2]. The Methods sections of the articles were reviewed, and the methodology categorized according to the Beaton, *et al* "cube" classification system for studies of responsiveness[3].

Nine procedures identified were classified as follows: Patient perspective I: comparison to a global rating[4-7]; Patient perspective II: patient conversation[8,9]; Clinician perspective I: consensus development (Delphi)[10-12]; Clinician perspective II: patient scenario scoring[13]; Clinician perspective III: patient scenario comparison[14]; Clinician perspective IV: prognostic rating scale[15]; Data driven approach[16-19]; Discerning important improvement I: improvement criteria[20,21]; Discerning important improvement II: achieving treatment goals[22].

Methods included that of Jaeschke, *et al*[4], known as the anchor-based method in which the relationship between an outcome measure of interest and an independent measure (or anchor) is examined to elucidate the meaning of a particular degree of change. The anchor approach is used here, but since an external anchor was not available, an internal anchor(s) was selected and changes in PRO related to MCID of the internal anchor are considered using procedures that mimic approaches used for external anchors.

Our goal was to determine for patients with RA the MCID in 3 significant patient-reported outcomes: activity limitation, fatigue, and sleep quality.

MATERIALS AND METHODS

Datasets. The data from 2 randomized, double-blind, placebo controlled trials in patients with active RA were used for our evaluation. The data were made available in SAS Version 9.1 export files (SAS Institute Inc., Cary, NC, USA).

The ATTAIN Study[23]. The first randomized controlled trial (RCT) was a Phase III multicenter, 6-month trial evaluating the efficacy and safety of abatacept on a background of DMARD therapy in patients with active RA (RA functional class I, II, or III) who were anti-tumor necrosis factor (TNF) therapy failures. Eligible and consenting patients were randomized 2:1 to receive abatacept (n = 258) or placebo (n = 133) on a background of disease modifying antirheumatic drugs (DMARD). The primary objectives were to compare abatacept to placebo regarding clinical efficacy as assessed by the American College of Rheumatology response criteria (ACR 20) response rate at 6 months; and the improvement in physical function as assessed by the Health Assessment Questionnaire (HAQ) disability index at 6 months. The primary efficacy outcome measures were: proportion of patients achieving an ACR 20 response at 6 months; the proportion of patients with physical function (HAQ) response as measured by improvement of at least 0.3 units from baseline in the HAQ disability index at 6 months.

The AIM Study[24]. The second RCT was a Phase III multicenter, 12-month trial evaluating efficacy and safety of abatacept on a background of methotrexate (MTX) therapy in patients with active RA (RA functional class I, II, or III) who had an inadequate response to MTX. Eligible and consenting patients were randomized 2:1 to receive abatacept (n = 433) or placebo (n = 219) on a background of MTX. The primary objectives were to compare abatacept to placebo regarding clinical signs and symptoms of RA as measured by ACR 20 response following 6 months of treatment; physical function by the HAQ disability index at 12 months; and radiographic progression by erosion score using the Genant-modified Sharp method at 12 months of treatment. The primary efficacy outcome measures were: proportion of patients achieving ACR 20 response at 6 months; proportion of patients with physical function (HAQ) response as measured by improvement of at least 0.3 units from baseline in the HAQ disability index at 12 months; and erosion score at 12 months using the Genant modified Sharp method.

Outcome measures. For both RCT a large number of outcome measures were assessed. The patient-reported outcomes (PRO) of interest in this assessment were activity limitation, fatigue, and sleep problems. A 2-item questionnaire was developed to collect data on amount of time a patient was unable to perform usual activities because of RA during the previous 30 days; a validated fatigue assessment measure was used to evaluate fatigue[25]; and the validated 12-item Medical Outcomes Study sleep questionnaire was used to measure sleep quality, with an overall problems index generated as a summary measure of the different types of sleep problems[26]. Activity score: 0-30 days of limitation (0 = best, 30 = worst); Fatigue score: 0–100 (0 = no fatigue, 100 = complete fatigue); Sleep score: 0–100 (0 = no problems, 100 = worst case). For all 3 of these PRO, the low score is better and a negative change from baseline (i.e., final score – baseline score) is an improvement. The internal anchors that will be considered in assessing the patient-reported outcomes of interest are: HAQ: 0–3 (0 = best, 3 = worst); Patient global: 0–100 (0 = best, 100 = worst); Pain: 0–100 (0 = best, 100 = worst). For all 3 of these internal anchors, the low score is better and a negative change from baseline (i.e., final score – baseline score) is an improvement.

Anchor-based approach for determining MCID. MCID anchor-based methods were used to examine the relationship between an outcome measure of interest and an independent measure (or anchor) to elucidate the meaning of a particular degree of change. The anchor-based approaches described in Wells, *et al*[2] used external assessment of overall rating of change in patients, and MCID in the outcome measure of interest was derived by evaluating the relationship between these global ratings and changes in the outcome measure. For the data presented here, an external assessment of overall rating of change was not available. The anchor based approach used a composite of procedures presented in Wells, *et al*[2]. Since an external assessment or anchor was not available, an internal anchor(s) was selected, and changes in the PRO about the MCID of the internal anchor were considered using procedures mimicking approaches used for the external anchors.

Two operational issues need to be addressed for internal anchors: first, selection of the neutral point; and second, the selection of the narrow range about the neutral point based on grades of "worse" and "better" about this point. The MCID was selected as the neutral point based on the premise that if MCID of the internal anchor was considered to be "the smallest change or difference" in the anchor that is perceived as beneficial and would lead to a change in the patient's medical management, then for those patients in a narrow range about this point the corresponding change or difference in the PRO of interest would represent a small but important change or difference. The rationale for the choice of the narrow range to be ± 1/4 standard deviation (SD) was based on 1/4 SD/SD = 0.25 corresponding to a small effect size. This satisfies the requirements for determining MCID for an outcome measure: an indicator that change has occurred or that a difference exists (based on the choice of the internal anchor); an important observed change or difference based on a valid assignment of importance (based on the MCID of the internal anchor); and an appropriate method to determine the threshold level within the distribution of important change or difference scores (based on the concept of "small effect size").

The specific steps to derive the MCID were as follows:

1. The MCID for the internal anchor(s) was selected. For patients with RA, Wells, *et al*[9] determined the point at which differences in clinical assessment scores on physical activity (HAQ), pain, and overall condition (patient global) were sufficiently large to correspond to a subjective perception of a meaningful difference from the perspective of the patient. The patients rated them-

selves as "somewhat better" than their conversational partner when they had a 7.2% better score on the HAQ (note: 7.2% difference on the HAQ corresponds to a difference of 0.22 units in the HAQ score), 6.2% less pain on a 10 cm visual analog scale (VAS), and 9.1% better global assessment on a 10 cm VAS. The MCID taken from this study were: 0.20 for the HAQ, 10 for pain on a 100 mm VAS, and 10 for patient global on a 100 mm VAS.

2. The orderly progression of the change in a PRO was evaluated over a range of values of the internal anchor given in steps of 0.5 SD above and below its MCID.

3. As for the external anchor approach of methods for determining a MCID, the change in the PRO was considered based on converging steps about a central point. In this case the central point was taken as the MCID of the internal anchor. The range of the central point was:

$$(\text{MCID} + 0.25\ \text{SD}, \text{MCID}) \text{ to } (\text{MCID} - 0.25\ \text{SD}, \text{MCID}).$$

4. Steps 2 and 3 were repeated for each internal anchor.

5. A Delphi approach was taken to reach consensus on the MCID based on the range of values derived for the internal anchors.

Five methodologists reviewed the results of steps 2 and 3 for each internal anchor according to a Delphi process. That is, along with a summary of the results, they were provided with a description of the activity, fatigue, and sleep scores (i.e., wording of the question and response key) and requested to provide their best estimate of the MCID for each of the 3 outcomes. Their responses were averaged, and this summary was sent back with an invitation to revise their estimate in light of the amalgamated responses. This process was continued until consensus was achieved. The methodologists included 2 physicians (a rheumatologist and general internist) with clinical epidemiology training and experience in clinical trials and evidence-based medicine; a biostatistician with experience in outcome measures in rheumatology and analysis of trial data; a clinical trial methodologist with experience in designing, conducting, and interpreting trials in RA; and a health technology assessor with experience in health economics studies related to arthritis trials.

The effect sizes for the MCID were calculated based on the ratio of the derived MCID and a weighted average of the corresponding SD.

RESULTS

The mean (SD) for the 3 PRO of interest (i.e., activity score, fatigue score, and sleep score) and the 3 internal anchors (i.e., HAQ, patient global, and pain) are displayed in Table 1. In general, all 3 PRO of interest improved between baseline and final assessments. For each study considered individually as well as combined, the change from baseline for activity limitation was negative, indicating improvement in patient activity. In particular, a 43% reduction in the number of days of limitation was observed when the 2 studies were combined. Similar negative change scores indicating improvement were found for fatigue and sleep, with a 31% reduction in the VAS fatigue score and 19% reduction in the VAS sleep index score between baseline and final assessment when both studies are combined. For each study considered individually as well as combined, the change from baseline for all 3 internal anchors showed improvement (Table 1). In particular, when both studies were combined, reductions of 39% reduction in patient assessment, 27% in HAQ, and 38% in pain were observed.

Activity limitation. A negative change from baseline assessment in the outcome measure "activity limitation" and the 3 internal anchors all correspond to improvement. A consistent change in activity limitation was found for changes in all 3 internal anchors: HAQ, patient global, and pain (Figure 1). In all situations a decreasing score for activity limitation was observed for decreasing scores of the 3 internal anchors. That is, activity limitation was better for improved levels of function, patient global assessment, and pain assessment. These same patterns held for each of the studies ATTAIN and AIM. For determining MCID, the change in the activity score was considered based on converging steps about the MCID of the internal anchor. In particular, changes within a narrow range

Table 1. Descriptive statistics — patient-reported outcomes and internal anchors.

	Assessment	ATTAIN Study	AIM Study	Both Studies
Patient-reported outcome				
Day's activities limited	Baseline	16.81 ± 11.12	13.74 ± 11.26	14.87 ± 11.30
	End of Study	11.35 ± 11.43	6.70 ± 10.01	8.41 ± 10.79
	Change	–5.32 ± 11.95	–7.09 ± 12.22	–6.44 ± 12.14
Subject's fatigue	Baseline	73.07 ± 19.51	63.88 ± 23.13	67.25 ± 22.31
	End of Study	56.56 ± 27.87	40.83 ± 27.73	46.59 ± 28.78
	Change	–16.57 ± 29.43	–23.09 ± 28.31	–20.71 ± 28.88
Sleep Problems Index II	Baseline	47.90 ± 18.83	43.42 ± 20.26	45.06 ± 19.86
	End of Study	40.69 ± 19.27	34.07 ± 19.15	36.50 ± 19.45
	Change	–7.18 ± 16.82	–9.35 ± 17.03	–8.55 ± 16.98
Internal anchor				
Patient's assessment	Baseline	69.10 ± 19.87	62.57 ± 21.39	64.96 ± 21.07
	End of Study	49.13 ± 27.80	34.54 ± 26.02	39.97 ± 27.60
	Change	–19.87 ± 29.65	–28.09 ± 28.84	–25.04 ± 29.40
HAQ	Baseline	1.83 ± 0.59	1.68 ± 0.64	1.74 ± 0.63
	End of Study	1.49 ± 0.72	1.12 ± 0.72	1.26 ± 0.74
	Change	–0.33 ± 0.55	–0.55 ± 0.66	–0.47 ± 0.63
Pain	Baseline	70.22 ± 19.42	64.21 ± 20.96	66.41 ± 20.60
	End of Study	49.79 ± 28.17	35.82 ± 26.90	41.02 ± 28.18
	Change	–20.44 ± 31.05	–28.30 ± 29.13	–25.39 ± 30.08

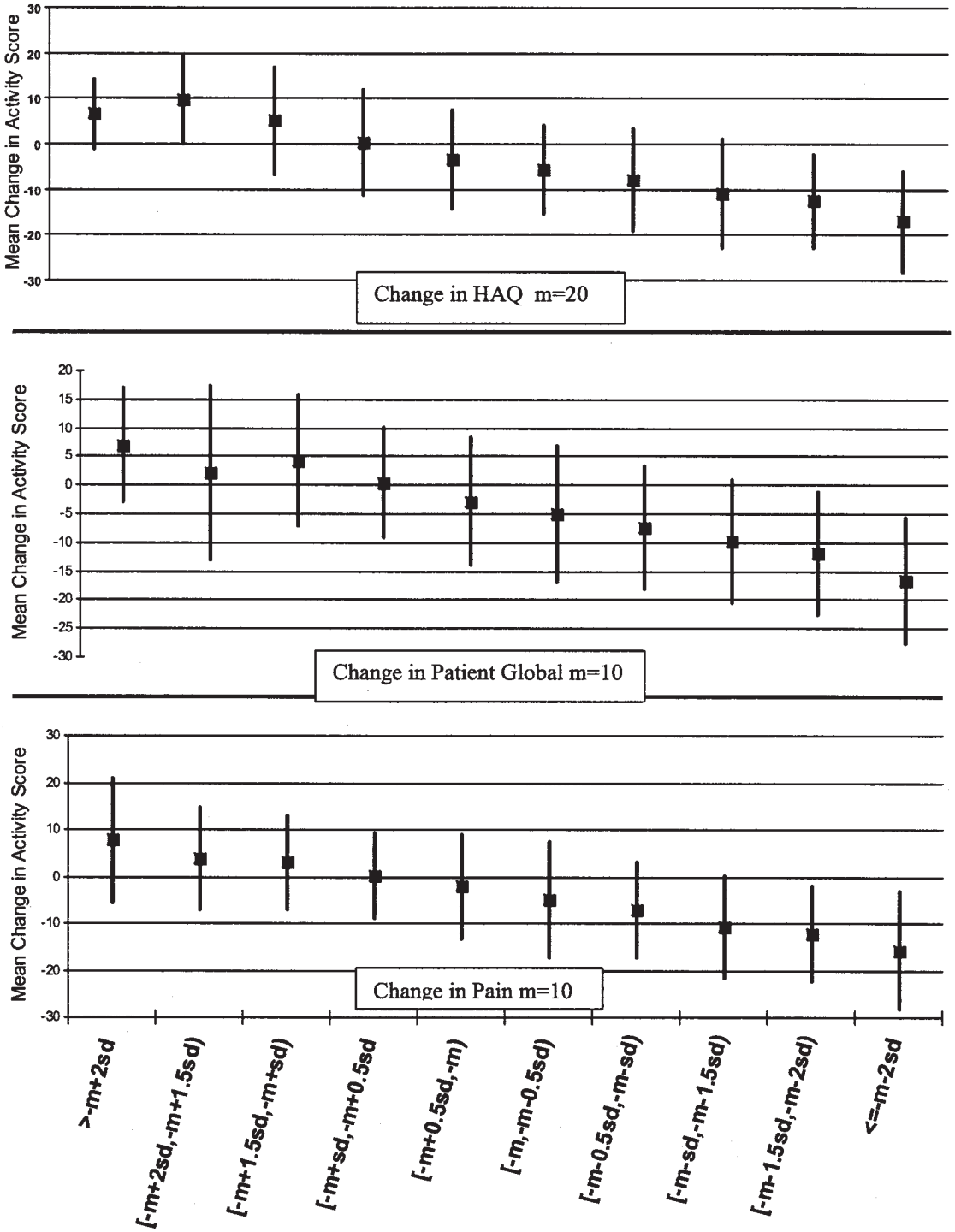

Figure 1. ATTAIN and AIM studies: Mean changes for Activity Score over a range of values of the internal anchors (mean ± standard deviation).

of ± 0.25 SD about the MCID of –0.20 for the HAQ, 10 for the patient global, and 10 for pain are considered (Table 2). After 2 rounds of the Delphi exercise with 5 methodologists, a consensus of 4 days was arrived at for the MCID for activity limitation.

Fatigue. As for activity limitation, a negative change from the baseline assessment in the outcome measure "fatigue" and the 3 internal anchors all correspond to improvement. A consistent change in fatigue was found for changes in all 3 internal anchors: HAQ, patient global, and pain (Figure 2). In all situations a decreasing score for fatigue was observed for decreasing scores of the 3 internal anchors. That is, fatigue was bet-

Table 2. Mean changes for Activity, Fatigue, and Sleep scores in a narrow range about the MCID of the internal anchors [mean ± standard deviation (number of patients)].

Internal Anchor	MCID of Anchor	ATTAIN	AIM	Both
Activity score				
HAQ	0.25 SD above	–4.4 ± 11.6 (44)	–3.4 ± 8.4 (46)	–3.9 ± 10.0 (90)
	0.25 SD below	–4.2 ± 9.8 (46)	–6.4 ± 9.7 (43)	–5.3 ± 9.8 (89)
Patient global	0.25 SD above	–4.3 ± 9.7 (36)	–3.4 ± 9.8 (54)	–3.7 ± 9.7 (91)
	0.25 SD below	–2.6 ± 9.5 (46)	–4.4 ± 14.8 (49)	–3.4 ± 12.5 (95)
Pain	0.25 SD above	–5.3 ± 7.8 (26)	–2.9 ± 10.1 (53)	–3.7 ± 9.4 (80)
	0.25 SD below	–2.8 ± 9.9 (46)	–4.7 ± 11.2 (57)	–3.9 ± 10.6 (103)
Fatigue score				
HAQ	0.25 SD above	–12.5 ± 27.0 (46)	–13.1 ± 25.1 (49)	–12.5 ± 25.8 (94)
	0.25 SD below	–9.7 ± 30.5 (47)	–19.0 ± 23.9 (45)	–14.3 ± 27.7 (92)
Patient global	0.25 SD above	–2.7 ± 18.5 (35)	–8.2 ± 17.1 (54)	–6.0 ± 17.7 (90)
	0.25 SD below	–9.7 ± 18.5 (46)	–13.8 ± 21.3 (49)	–11.8 ± 20.0 (95)
Pain	0.25 SD above	–5.8 ± 16.8 (25)	–8.0 ± 20.6 (54)	–7.3 ± 19.3 (80)
	0.25 SD below	–9.6 ± 18.8 (46)	–15.4 ± 18.5 (57)	–12.8 ± 18.8 (103)
Sleep score				
HAQ	0.25 SD above	–8.6 ± 14.0 (46)	–4.8 ± 12.4 (49)	–6.7 ± 13.4 (94)
	0.25 SD below	–6.8 ± 15.7 (47)	–5.5 ± 15.2 (45)	–6.2 ± 15.4 (92)
Patient global	0.25 SD above	–7.8 ± 16.6 (37)	–4.1 ± 13.1 (54)	–5.6 ± 14.6 (92)
	0.25 SD below	–4.4 ± 12.6 (46)	–4.1 ± 16.5 (49)	–4.2 ± 14.7 (95)
Pain	0.25 SD above	–5.3 ± 15.5 (26)	–2.8 ± 16.3 (54)	–3.8 ± 16.0 (81)
	0.25 SD below	–5.7 ± 16.3 (47)	–7.1 ± 16.0 (57)	–6.4 ± 16.0 (104)

ter for improved levels of function, patient global assessment, and pain assessment. These same patterns held for each of the studies ATTAIN and AIM. For determining MCID, the change in the fatigue score was considered based on converging steps about the MCID of the internal anchor. In particular, changes within a narrow range of ± 0.25 SD about the MCID of –0.20 for the HAQ, 10 for the patient global, and 10 for pain are considered (Table 2). After 3 rounds of the Delphi exercise with 5 methodologists, a consensus of 10 points on the 100 VAS scale was arrived at for the MCID for fatigue.

Sleep quality. A negative change from the baseline assessment in the outcome measure "sleep" and the 3 internal anchors all correspond to improvement. A consistent change in sleep was found for changes in all 3 internal anchors: HAQ, patient global, and pain (Figure 3). In all situations a decreasing score for sleep was observed for decreasing scores of the 3 internal anchors. That is, sleep was better for improved levels of function, patient global assessment, and pain assessment. These same patterns held for each of the studies ATTAIN and AIM. For determining MCID, the change in the sleep score was considered based on converging steps about the MCID of the internal anchor. In particular, changes within a narrow range of ± 0.25 SD about the MCID of –0.20 for the HAQ, 10 for the patient global, and 10 for pain are considered (Table 2). After 3 rounds of the Delphi exercise with 5 methodologists, a consensus of 6 points on the 100 VAS scale was arrived at for the MCID for sleep.

Consistent patterns of change in the 3 patient-reported outcomes and the internal anchors were found for both studies, with correlations in the range of 0.5, 0.7, and 0.4 for activity, fatigue, and sleep, respectively. Following the procedure for using internal anchors described in Materials and Methods, the minimal clinically important differences for activity limitation, fatigue, and sleep problems were determined (Table 3). The mean changes for activity, fatigue, and sleep in a narrow range about the MCID of the 3 internal anchors corresponding to the 2 trials were found, and following the Delphi process the MCID determined were 4 for activity limitation, 10 for fatigue, and 6 for sleep problems. The corresponding effect sizes were 0.39, 0.46, and 0.40 for activity limitation, fatigue, and sleep problems, respectively.

A sensitivity analysis based on the selection of the MCID for HAQ was conducted. In particular, when an MCID of –0.25 was selected for the HAQ, the resulting mean changes in a narrow range about the MCID for the HAQ for the patient-reported outcomes were of a similar magnitude and, as a result, the MCID for these patient-reported outcomes would be similar. For activity, fatigue, and sleep, respectively, the mean changes were 4.6, 12.8, and 4.2 for HAQ of –0.25 (compared to 4.3, 10.3, 7.1 for HAQ of –0.20) in the AIM study, and 4.8, 14.8, and 7.3 for a HAQ of –0.25 (compared to 4.3, 17.0, 7.3 for HAQ of –0.20) for the ATTAIN study.

DISCUSSION

MCID anchor-based methods examine the relationship between an outcome measure of interest and an independent measure (or anchor) to elucidate the meaning of a particular degree of change. After several years of development and

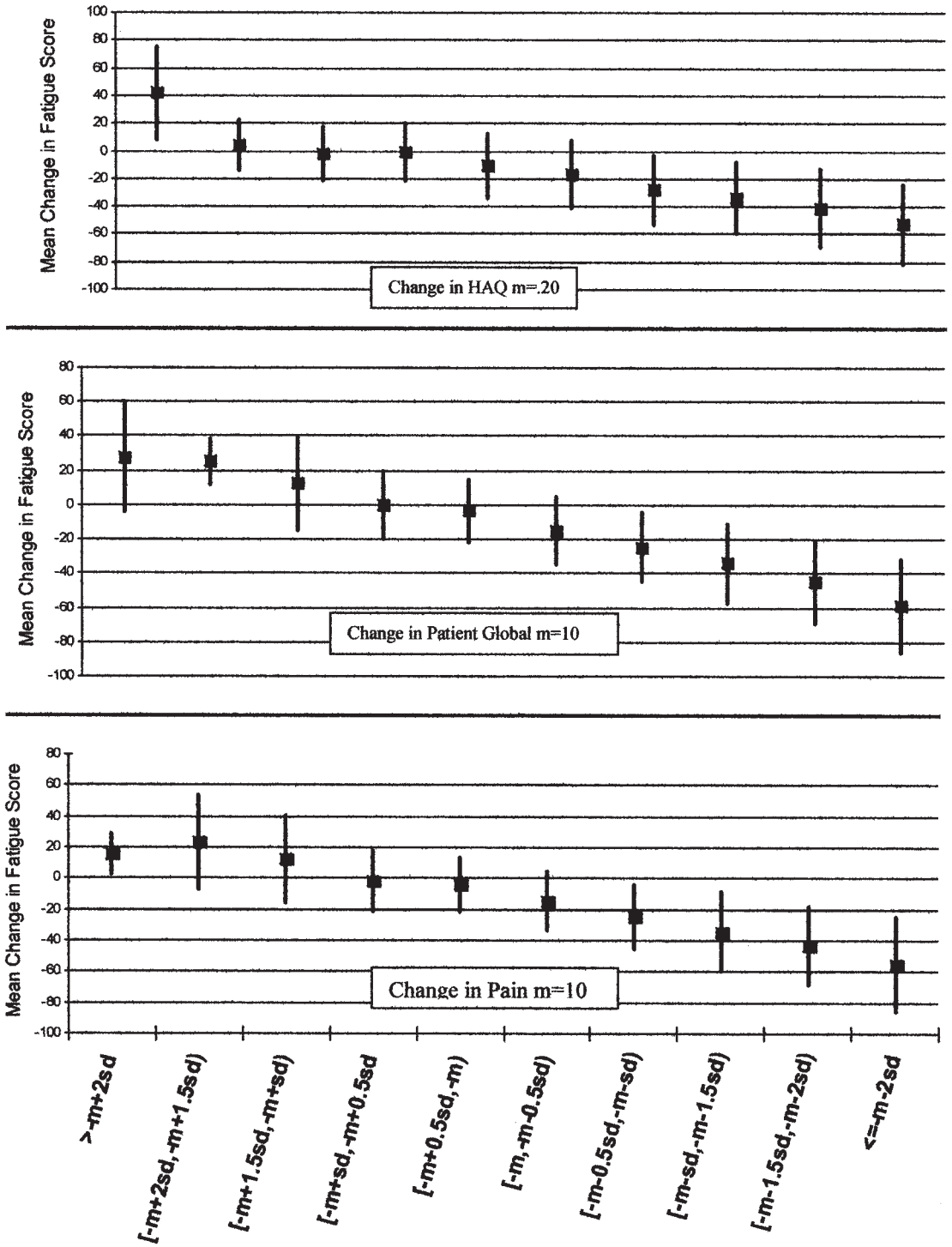

Figure 2. ATTAIN and AIM studies: Mean changes for Fatigue Score over a range of values of the internal anchors (mean ± standard deviation).

implementation, acceptance of determining MCID using external anchors is established. When MCID are derived based on existing data, internal anchors remain the only recourse in an anchor-based approach.

The internal anchor approach is well founded, since it is based on a composite of features of procedures for determining MCID that are available in the literature[2]. As for the external anchor methods (see above: Patient perspective I: Comparison to a global rating; Patient perspective II: Patient conversation; Clinician perspective III: Patient scenario com-

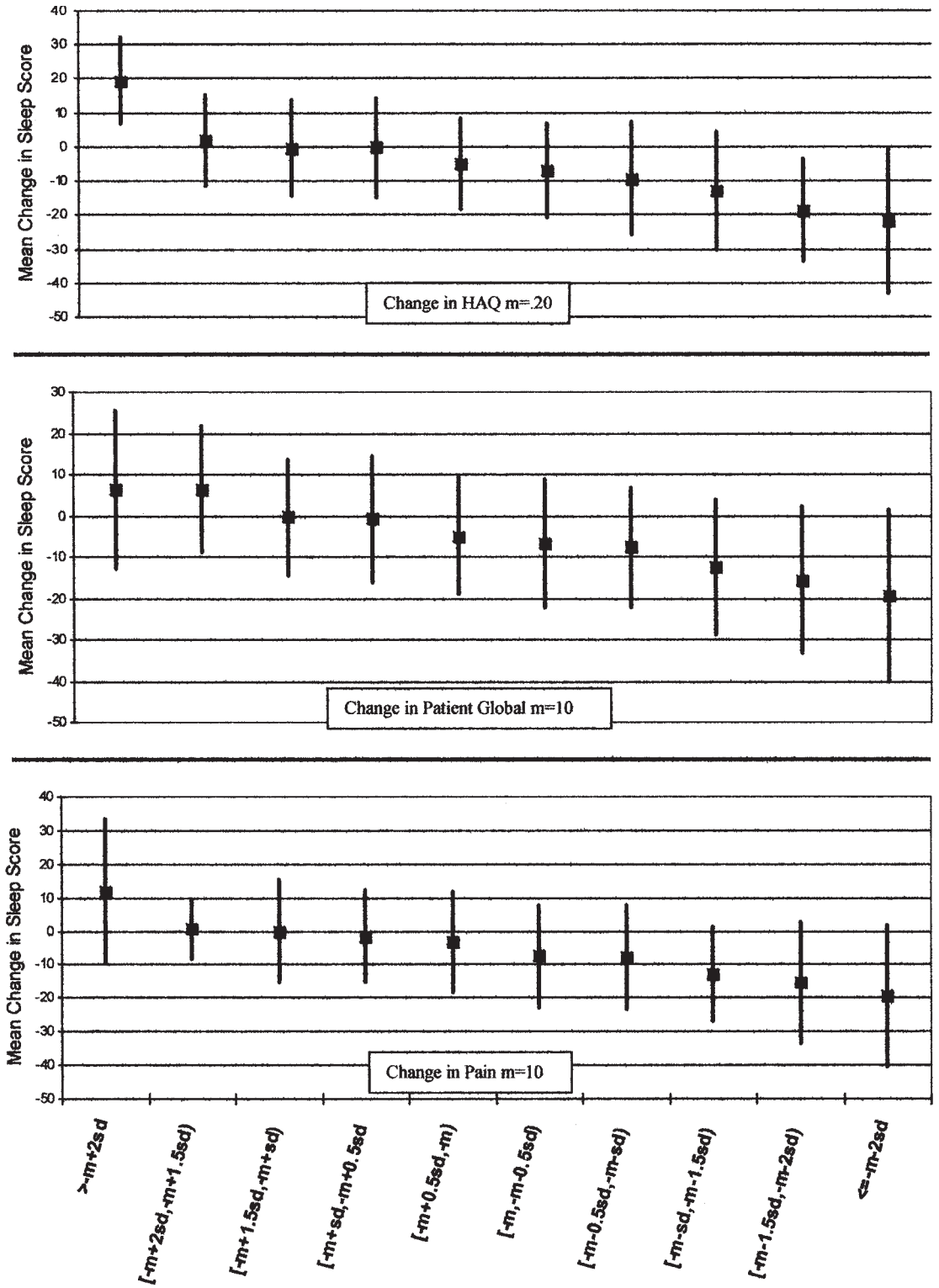

Figure 3. ATTAIN and AIM studies: Mean changes for Sleep Score over a range of values of the internal anchors (mean ± standard deviation).

parison; Clinician perspective IV: Prognostic rating scale; and Discerning important improvement II: Achieving treatment goals, a change in the PRO will be considered based on converging steps about a central point. In this case the central point is taken as the MCID of the internal anchor. Next, as for the Clinician perspective I: Consensus development (Delphi)

Table 3. Mean changes for Activity, Fatigue, and Sleep scores in a narrow range about the MCID of the internal anchors.

Outcome	6-month Study (ATTAIN)			12-month Study (AIM)			MCID
	Anchor HAQ	Anchor Patient Global	Anchor Pain	Anchor HAQ	Anchor Patient Global	Anchor Pain	
Activity (0–30)	4.3	3.4	3.7	4.3	3.8	3.9	4
Fatigue (0–100)	10.3	6.7	8.3	17.0	10.9	11.8	10
Sleep (0–100)	7.1	5.9	5.5	7.3	4.1	5.0	6

method for determining a MCID, a Delphi approach will be taken to reach consensus on the MCID based on the range of values derived for the internal anchors. Finally, similar to the Discerning important improvement I: Improvement criteria method for determining MCID, the MCID definitions derived will be assessed in data sets of appropriate placebo-controlled trials with interventions that offer the largest possible efficacy difference between intervention and placebo and include the outcome measures of interest.

Others have considered a similar conceptual basis for this procedure. For example, in the study by Kosinski, *et al*[27], all patients enrolled in 2 double-blind, placebo-controlled clinical trials (n = 693) designed to assess the efficacy and safety of RA treatment were grouped together and considered in the evaluation. The minimally important changes in Medical Outcome Study Short Form-36 (SF-36) and HAQ scores were determined in 2 ways: first, by estimating the mean changes of these scores for patients who showed one level of improvement on the 5 RA severity measures; and second, by determining the percentage of patients whose followup scores improved over baseline scores more than would be expected by chance for patients who showed one level of improvement on the 5 RA severity measures. However, the conceptual and empirical basis for these levels and categories of the core set measures and the importance of a one-level change in any of the measures were not provided.

The choice of the 3 internal anchors for determining the MCID was based on practical and conceptual issues. The patient global assessment of disease was selected, as it most closely aligns itself with the type of anchor used in establishing MCID in which an overall assessment is made. The baseline values are then used to establish change, and the MCID of the patient global is taken as a small but important change in the anchor. Since patient global was not designed specifically as an anchor for establishing MCID, other frames of reference for change were selected, and consistency among the several perspectives was assessed. In choosing the other anchors, a functional ability and a physical variable were considered appropriate choices: HAQ and pain assessment were taken, since the MCID for these 2 core set measures were available. The conceptual relationship between these 2 anchors and activity, fatigue, and sleep is based on our related work in assessing these outcomes in relationship to the domains and component scores of the SF-36. We found that anchors HAQ and pain assessment were related to the physical component of the SF-36; activity limitation was related more to the physical component; fatigue was related to both physical and mental components; and sleep was more closely related to the mental component. When physical variables such as pain and HAQ improve, then activity limitation and fatigue improve directly, while sleep improves indirectly through its relationship with other measures such as fatigue. The MCID determined for sleep was small: this could be rooted in the above relationship of sleep to these anchors; moreover, sleep has been rated as very important by patients, so even small changes are considered important. Further work is required to elucidate this relationship and relative importance.

Based on our approach described above we determined the following MCID: 4 for activity limitation, 10 for fatigue, and 6 for sleep problems. These differences can be used in the design of clinical trials focusing on these patient-reported outcomes and may provide a benchmark in assessing patient improvement.

Application of MCID: responder analysis. These MCID can be used as benchmarks in a responder analysis assessing patient improvement on these patient-reported outcomes, as follows: the value of each patient's PRO is compared to the MCID for that outcome; if the value exceeds the MCID, then the patient is considered to be responsive for that outcome; and the proportion for the "treatment" group that are responsive is compared to the proportion for the "comparison" group. A similar approach was taken for the ACR 20, ACR 50, and ACR 70 criteria.

As an illustration, consider the ATTAIN study[23], a Phase III multicenter, 6-month RCT evaluating the efficacy and safety of abatacept on a background of DMARD therapy in patients with active RA who were anti-TNF therapy failures. Of the patients randomized to receive abatacept, 53%, 59%, and 69% met or exceeded the MCID for the activity, fatigue, and sleep scores, respectively, compared to 31%, 37%, and 38% for patients receiving placebo. For all 3 patient-reported outcomes (activity, fatigue, sleep), the difference between abatacept and placebo was statistically significant (Table 4). Similarly, for the AIM study[24], a Phase III multicenter, 12-month RCT evaluating efficacy and safety of abatacept on a background of MTX therapy in patients with active RA with inadequate response to MTX, 59%, 69%, and 58% of the patients receiving abatacept plus MTX met or exceeded the

Table 4. Responder analysis: percentage of patients exceeding the MCID of the patient-reported outcome.

Outcome	6-month Study (ATTAIN)			12-month Study (AIM)		
	Abatacept	Placebo	p	Abatacept + MTX	Placebo + MTX	p
Activity, MCID 4 (%)	138 (53.1)	41 (31.1)	< 0.0001	264 (58.7)	101 (44.5)	0.0007
Fatigue, MCID 10 (%)	152 (58.5)	49 (37.1)	< 0.0001	311 (69.1)	115 (51.1)	< 0.0001
Sleep, MCID 6 (%)	153 (58.9)	50 (37.9)	< 0.0001	261 (58.0)	105 (46.7)	0.0052

MCID for activity, fatigue, and sleep scores, respectively, compared to 45%, 51%, and 47% for patients receiving placebo plus MTX. The difference between abatacept plus MTX and placebo plus MTX was statistically significant for each of the patient-reported outcomes (Table 4).

Application of the MCID. Clinical importance descriptors. The MCID for activity, fatigue, and sleep can help describe and interpret change in the patient-reported outcomes from baseline. Comparing the mean change from baseline in a patient-reported outcome in the treatment group versus the comparison group, the p value provides an indication of the statistical significance. The MCID will provide an indication of whether the mean change found was clinically important (i.e., mean change exceeds MCID). In this way, for each treatment group the clinical importance of the change relative to MCID can be viewed. Further, for each study group, the 95% CI for the mean change from baseline can be calculated for assessing the significance of the clinical importance. If the lower confidence limit (LCL) of the 95% CI exceeds the MCID, then the clinical importance found is considered significant.

As an illustration, consider the ATTAIN study. Of the patients randomized to receive abatacept, a mean change from baseline of –7.4, –22.3, and –9.8 was found for activity, fatigue, and sleep, respectively, compared to –1.3, –5.3, –2.1 for patients receiving placebo. For all 3 patient-reported outcomes, the difference between abatacept and placebo was statistically significant (Table 5). In assessing the clinical importance of change in patient-reported outcomes, it is noted that for the placebo group no change from baseline for activity, fatigue, and sleep exceeded the corresponding MCID, whereas for the abatacept group all change exceeded the MCID. Further, for the abatacept group, the lower confidence interval for all 3 patient-reported outcomes exceeded the corresponding MCID, indicating that the clinical importance is significant. For the AIM study, a mean change from baseline for activity, fatigue, and sleep of –8.4, –25.9, and –10.4 was found for patients receiving abatacept + MTX, compared to –4.5, –17.3, and –7.2 for patients receiving placebo + MTX. For all 3 patient-reported outcomes (activity, fatigue, and sleep), the difference between abatacept and placebo was statistically significant (Table 5). Changes for all the patient-reported outcomes for both study groups exceeded the corresponding MCID, but for abatacept + MTX the lower confidence interval for all 3 patient-reported outcomes exceeded the corresponding MCID, indicating significance of clinical importance, whereas for the placebo + MTX group the LCL exceeded the MCID only for fatigue.

Having derived MCID for patient-reported outcomes of activity, fatigue, and sleep, the next step is to assess the derived definitions in other data sets of appropriate placebo-controlled trials with interventions that offer the largest possible efficacy difference between intervention and placebo and that include outcome measures of interest.

Table 5. Mean change from baseline of the patient-reported outcomes.

Outcome	6-month Study (ATTAIN)			12-month Study (AIM)		
	Abatacept Δ (95% CI)	Placebo Δ (95% CI)	p	Abatacept + MTX Δ (95% CI)	Placebo + MTX Δ (95% CI)	p
Activity (MCID 4)	–7.4 (–8.9, –5.9)	–1.3 (–3.3, 0.6)	< 0.0001	–8.4 (–9.4, –7.3)	–4.5 (–6.3, –2.7)	0.0002
Fatigue (MCID 10)	–22.3 (–25.8, –18.7)	–5.3 (–10.1, –0.6)	< 0.0001	–25.9 (–28.6, –23.3)	–17.3 (–21.0, –13.7)	0.0003
Sleep (MCID 6)	–9.8 (–11.8, –7.7)	–2.1 (–4.8, 0.7)	< 0.0001	–10.4 (–12.0, –8.7)	–7.2 (–9.3, –5.2)	0.0187

Δ: mean change from baseline.

REFERENCES

1. Wells GA, Anderson J, Boers M, et al. MCID/Low Disease Activity State Workshop summary recommendations and research agenda. J Rheumatol 2003;30:1115-8.
2. Wells GA, Beaton D, Shea B, et al. Minimal clinically important differences: review of methods. J Rheumatol 2001;28:406-12.
3. Beaton DE, Bombardier C, Katz JN, et al. Looking for important changes/differences in studies of responsiveness. J Rheumatol 2001;28:400-5.
4. Jaeschke R, Singer J, Guyatt GH. Measurement of health status ascertaining the minimal clinically important difference. Control Clin Trials 1989;10:407-5.
5. Juniper EF, Guyatt GH, Willan A, Griffith L, determining a minimal important change in a disease-specific quality of life questionnaire. J Clin Epidemiol 1994;47:81-7.
6. Stratford PW, Binkley JM. Measurement properties of the RM-18: a modified version of the Roland-Morris Disability Scale. Spine 1997;22:2416-21.
7. Stratford PW, Binkley JM, Riddle DL, Guyatt GH. Sensitivity to change of the Roland-Morris Back Pain Questionnaire. 1. Physical Therapy 1998;78:1186-96.
8. Redelmeier DA, Lorig K. Assessing the clinical importance of symptomatic improvements: an illustration in rheumatology. Arch Intern Med 1993;153:1337-42.
9. Wells GA, Tugwell P, Kraag GR, Baker PRA, Groh J, Redelmeier DA. Minimum important difference between patients with rheumatoid arthritis: the patient's perspective. J Rheumatol 1993;20:557-60.
10. Bellamy N, Carette S, Ford PM, et al. Osteoarthritis antirheumatic drug trials. III. Setting the delta for clinical trials — results of a consensus development (Delphi) exercise. J Rheumatol 1992;19:451-7.
11. Bellamy N, Anastassiades TP, Buchanan WW, et al. Rheumatoid arthritis antirheumatic drug trials. III. Setting the delta for clinical trials of antirheumatic drugs — results of a consensus development (Delphi) exercise. J Rheumatol 1991;18:1908-5.
12. Bellamy N, Buchanan WW, Esdaile JM, et al. Ankylosing spondylitis antirheumatic drug trials. III. Setting the delta for clinical trials of antirheumatic drugs — results of a consensus development (Delphi) exercise. J Rheumatol 1991;18:1716-22.
13. van Walraven C, Mahon J, Moher D, Bohm C, Laupacis A. Surveying physicians to determine the minimal important difference: implications for sample size calcuation. J Clin Epidemiol 1999;52:717-23.
14. Todd KH, Funk JP. The minimum clinically important difference in physician-assigned visual analog pain scores. Acad Emerg Med 1996;3:142-6.
15. Stratford PW, Riddle DL, Binkley JM, Spadoni G, Westaway MD, Padfield B. Using the neck disability index to make decisions concerning individual patients. Physiotherapy Canada 1999;Spring:107-12.
16. Wyrwich KW, Nienaber MA, 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.
17. Wyrwich KW, Tierney WM, Wolinsky FD. Further evidence supporting standard error of measurement based criterion for identifying meaningful intra-individual change in health-related quality of life. J Clin Epidemiol 1999;52:861-73.
18. Wyrwich KW, Fihn SD, Tierney WM, Kroenke K, Babu AN, Wolinsky FD. Clinically important changes in health-related quality of life for patients with chronic obstructive pulmonary disease. J Gen Intern Med 2003;18:196-202.
19. Wyrwich KW, Tierney WM, Wolinsky FD. Using the standard error of measurement to identify important changes on the Asthma Quality of Life Questionnaire. Qual Life Res 2002;11:1-7.
20. Felson DT, Anderson JJ, Boers M, et al. American College of Rheumatology preliminary definition of improvement in rheumatoid arthritis. Arthritis Rheum 1995;38:727-35.
21. Felson DT, Anderson JJ, Lange MLM, Wells GA, 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.
22. Riddle DL, Stratford PW, Binkley JM. Sensitivity to change of the Roland-Morris Back Pain Questionnaire: 2. Physical Therapy 1998;78:1197-207.
23. Genovese MC, Becker JC, Schiff M, et al. Abatacept for rheumatoid arthritis refractory to tumor necrosis factor alpha inhibition. N Engl J Med 2005;353:1114-23.
24. Kremer JM, Genant HK, Moreland LW, et al. Effects of abatacept in patients with methotrexate-resistant active rheumatoid arthritis: a randomized trial. Ann Intern Med 2006;144:865-76.
25. Wolfe F. Fatigue assessments in rheumatoid arthritis: comparative performance of VAS and longer fatigue questionnaires in 7760 patients. J Rheumatol 2004;31:1896-902.
26. Stewart AL, Ware AE. Measuring functioning and well being: The Medical Outcomes Study approach. Durham, NC: Duke University Press; 1992.
27. 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.