Establishment of the Minimum Clinically Important Difference for the Bath Ankylosing Spondylitis Indices: A Prospective Study

STEPHAN PAVY, SINEAD BROPHY, and ANDREI CALIN

ABSTRACT. ***Objective.*** The minimum clinically important difference (MCID) is the minimum level of change of an outcome measure that is considered to be clinically relevant. This prospective study was conducted to establish the minimum level of change in the Bath Ankylosing Spondylitis (BAS) indices that is meaningful for both the patient and the clinician.

Methods. The BAS questionnaires [i.e., Bath Ankylosing Spondylitis Functional Index (BASFI); Bath Ankylosing Spondylitis Disease Activity Index (BASDAI); and Bath Ankylosing Spondylitis patient Global score (BAS-G)] were administered to 125 patients with ankylosing spondylitis (AS) at baseline and at the end of a 2-week intensive physiotherapy program. Together with the final assessment, a global validated rating scale was used to examine each domain. Receiver operating characteristic (ROC) curves were plotted to determine the BAS change score that most accurately classified patients with respect to a clinically meaningful change. This analysis was repeated to investigate whether the estimate of change was dependent upon the patients' baseline scores.

Results. According to analyses of ROC curves, the MCID are 7 mm or 17.5% for BASFI: sensitivity = 0.60/specificity = 0.85; 10 mm or 22.5% for BASDAI: sensitivity = 0.65/specificity = 0.82; and 15 mm or 27.5% for BAS-G: sensitivity = 0.61/specificity = 0.74. MCID values were independent of the patients' baseline scores (p = 0.09 to 0.72) by regression analysis.

Conclusion. This prospective study gives MCID values for BASFI, BASDAI, and BAS-G that allow translation from the BAS indices to readily understood values for both patients and clinicians. (J Rheumatol 2005;32:80–5)

Key Indexing Terms:
ANKYLOSING SPONDYLITIS
MINIMUM CLINICALLY IMPORTANT DIFFERENCE
OUTCOME MEASURES
PATIENT PERSPECTIVE

Ankylosing spondylitis (AS) is a chronic inflammatory condition that is difficult to monitor because objective measures such as acute phase reactants fail to reflect disease activity[1]. In the past, the evaluation of chronic disease was often hampered by the lack of clearly defined outcome variables. With the development of the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI)[2], Bath Ankylosing Spondylitis Functional Index (BASFI)[3], and Bath Ankylosing Spondylitis patient Global score (BAS-G)[4] outcomes assessments as measures of disease activity, function, and patient global status, respectively, the relevance of these measurements needs to be established in both patient and clinical terms rather than less applicable statistical and numerical terms. To date, the BAS indices have commonly been used in placebo-controlled studies and clinical practice. Moreover, they have been endorsed by the Assessments in Ankylosing Spondylitis (ASAS) and Outcome Measures in Rheumatology (OMERACT) groups[5].

To interpret the result of a therapeutic intervention within the framework of a clinical trial or in daily practice, the clinician must translate a change of an outcome measure into a readily understood concept. Thus, one should be able to distinguish a clinically relevant difference, a statistically significant difference, or an irrelevant difference due to the error of measure. When the condition status is measured on a continuous scale, one should determine whether the observed difference establishes a minor or important effect on the health of the patient. To address this, the concept of "minimum clinically important difference" (MCID) was defined by Jaeschke, *et al*[6] as "the smallest difference in score in the domain of interest which patients perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive costs, a change in the patient's management."

Further, determination of MCID values allows (1) calculation of the sample size used in clinical trials; (2) expression of results of trials in a simplified way (e.g., the number

From Hôpital Cochin, Paris, France; The Clinical School, Swansea University, Swansea, Wales; Royal National Hospital for Rheumatic Diseases, Bath, England.

Supported by a financial grant from Centocor BV and Schering-Plough Ltd.

S. Pavy, MD, Cochin Hospital and Paris 5 University, Paris; S. Brophy, BSc, PhD, Clinical School, Swansea University; A. Calin MD, FRCP, (retired) Royal National Hospital for Rheumatic Diseases.

Address reprint requests to Dr. S. Pavy, Service de Rhumatologie A, Hôpital Cochin, 27 Rue du Faubourg Saint-Jacques, 75679 Paris cedex 14, France. E-mail: stephan.pavy@cch.ap-hop-paris.fr

Submitted February 25, 2004; revision accepted August 17, 2004.

of patients having reached a clinically relevant level of improvement); and (3) interpretation of the magnitude of the results.

Several methods have been used to determine the MCID according to outcome measure[7]. Three approaches can be distinguished: (1) the patient's perspective, (2) the clinician's/expert's perspective, and (3) data-driven measures. Further, the domain of validity of the MCID value depends on the methods used: the type of change (difference between groups/patients over time or change within a group) and level of application (group or patient). Regardless of the method used, 3 components are necessary: (1) an indicator that a change has occurred or that a difference exists; (2) the capacity of this indicator to determine a difference or a change considered without ambiguity as clinically important; and (3) an appropriate method to determine the threshold level within the distribution of important change or difference scores.

The OMERACT conference in May 2000 focused on the MCID of outcome in rheumatology[8]. MCID values have already been determined in low back pain[9], rheumatoid arthritis[10], osteoporosis[11], and osteoarthritis[12]. At this conference, it was recommended that MCID values are needed in the other rheumatic diseases such as AS and should focus on the patient's perspective in developing response criteria. The purpose of this prospective study was to develop and define the MCID values according to the patient's perspective for 3 outcome measures commonly used in AS: BASFI, BASDAI, and BAS-G.

MATERIALS AND METHODS

Each year, the Royal National Hospital for Rheumatic Disease (RNHRD) organizes 24 inpatient programs of intensive physiotherapy and hydrotherapy together with focus groups and education for AS patients. Each session lasts 2 weeks and includes an average of 10 patients. All patients are referred by a rheumatologist who determines that the deterioration in clinical status is such that the patient would benefit from intensive inpatient treatment or strategies to optimize education and self-support.

In this prospective study, which was approved by the local ethics committee and performed according to the rules and regulations of Good Clinical Practice, 152 patients aged older than 18 years were consecutively recruited between March and October 2001. The diagnosis of AS was confirmed according to the New York modified criteria[13] for all enrolled patients. Subsequently, 19 patients were excluded from the study. Some of these patients did not achieve the diagnostic criteria (n = 5), others were unable to read English (n = 3), and the remaining patients had other conditions associated with AS that may have adversely affected their disability (n = 11). Eight additional patients were lost to followup because they left the program prematurely. Thus, data for 125 patients were included in the analyses. For most patients, optimum drug therapy was established before enrollment, and patients were asked to stay on a stable dose of their standard medication during the 2-week program.

The objectives of the questionnaires were explained to each patient at enrollment and the final evaluation. Outcome measures of disease activity, function, and global well being were defined with BASDAI, BASFI, and BAS-G on the first and the last days of the program. BASDAI, BASFI, and BAS-G are self-administered instruments using a 100 mm visual analog scale (VAS) to assess AS symptoms during the previous week. In conjunction with the final assessment, a 15 point global rating scale, validated by Jaeschke, *et al*[6], was used to examine separately each domain of BASDAI, BASFI, and BAS-G.

Concerning the BASDAI, for example, patients were asked: "in terms of your disease activity [i.e., pain, fatigue, stiffness, tenderness) are you: the same (score 0); worse [i.e., hardly, a little, somewhat, moderately, good deal, great deal, very great deal "worse" (score –1 to –7)]; or better [hardly, a little, somewhat, moderately, good deal, great deal, very great deal "better" (score +1 to +7)]."

Descriptive statistics (number of patients, mean, standard deviation, minimum and maximum values) were used to summarize demographic characteristics (age, disease duration, age at onset) and baseline outcomes measures (BASFI, BASDAI, and BAS-G). To investigate whether the estimate of change was dependent on a patient's initial score, or demographic and severity variables such as disease duration, age of onset, hip involvement, iritis, regression techniques were used (Statview 4.5 F software). A distribution of patients according to the 15 point global rating scale and mean changes [absolute change and relative (%) change] in BAS score corresponding to the 15 point global rating scale is presented.

The MCID for each score was determined as the value (absolute and relative) that best discriminates between patients who experienced an important change from those who did not. According to previous studies using the same 15 point global rating scale, an important change was defined as a very great deal (+7), great deal (+6), or good deal better (+5) perception by the patient after the intervention. Receiver operating characteristic (ROC) curves were analyzed to identify MCID values[14]. A ROC curve was produced by plotting sensitivity (y-axis) against 1 minus specificity (x-axis) for as many values of change as possible from 0 to 100 mm. Pictorially, the MCID is the value of the BAS indices change score corresponding to the value closest to the top left-hand corner of the curve. Concordance rates between each specific index and the associated global scale were calculated[14].

RESULTS

This study enrolled 133 patients. The ratio of men to women who participated in the 2 week program was 3.3 to 1. Eight patients discontinued prior to the final evaluation because they developed other conditions (n = 3) or because they decided to leave the program for personal reasons (n = 5). Data for 125 AS patients were available for analysis (Table 1).

The wide range of disease duration and disease severity showed that patients who participated in the study represented the entire spectrum of disease, from mild to severe AS. Changes in BAS scores during the program are shown in Table 2.

Overall, the mean improvements over the 2 week program for BASDAI, BASFI, and BAS-G were 19%, 20%,

Table 1. Baseline characteristics of analyzed patients (n = 125).

	Mean (SD)	Minimum	Maximum
Age, yrs	43.6 (± 10.4)	18	67
Disease duration, yrs	20.7 (± 9.7)	2	40
Age of onset, yrs	22.4 (± 6.8)	11	45
BASFI initial score, mm	48.3 (± 18.7)	11.8	91.5
BASDAI initial score, mm	50.1 (± 18.3)	15.6	93.9
BAS-G initial score, mm	47.3 (± 21.1)	5	88
Percentage of patients with hip involvement	42.4		
Percentage of patients who experienced iritis	12.4		

Table 2. Distribution of changes in the BAS indices from baseline to Week 2 according to the 15 point global rating scale (Jaeschke[6]).

Outcome Measure, 0–100 mm (n = 125)	Important Change							No Change	Worse (–1/–7)
	Very Great Deal Better (+7)	Great Deal Better (+6)	Good Deal Better (+5)	Moderately Better (+4)	Somewhat Better (+3)	Little Better (+2)	Hardly Better (+1)		
BASFI, n	4	27	42	23	18	7	0	2	2
Δ BASFI, median	14	16	8.6	8.3	1.5	1.5		–4	–6.1
Δ BASFI, mean	14.6	14	9.4	8.2	1.4	1.9		–4	–6.1
% Change, median	47	29	20	16	3	2		–8	-24
BASDAI, n	4	19	46	19	14	9	2	4	8
Δ BASDAI, median	38.7	20.5	11.5	8.7	64	3.1	–7	2.5	2.7
Δ BASDAI, mean	39	17.2	13.4	9.3	4.9	0.7	–7	4.9	–1.3
% Change, median	70	41	25	14	11	3	–22	11	6
BAS-G, n	6	21	53	18	16	4	0	4	3
Δ BAS-G, median	28	28	14	14	11	12		–13	–9
Δ BAS-G, mean	30.8	25.1	15.8	10.8	8.2	5.3		–16.5	–11.2
% Change, median	67	49	31	21	23	20		–37	–86

and 15%, respectively, in terms of relative value, or 11.5 mm, 8.6 mm, and 14.5 mm, respectively, in terms of absolute value. However, in 20% of patients, the post-program BASDAI score was greater than the pre-program BASDAI score (mean = –7.2 mm). Similarly, 10% of patients deteriorated with respect to BASFI (mean = –6.2 mm), and 22% of patients reported higher BAS-G (mean = –11.2 mm). Only one patient scored worse for all 3 indices.

Using the 15 point global rating scale associated with each index, 5 patients reported worsened function or no improvement after the 2 week program. In terms of disease activity and global well being, 17 and 14 patients, respectively, did not experience improvement. Concordance rates between each specific index and the associated global scale were 0.78, 0.92, and 0.80 for disease activity, function, and global well being, respectively.

Distribution of patients and changes in BAS score relating to the 15 point global rating scale are shown in Table 2. Initial patient BAS scores were quite similar irrespective of the degree of global improvement (regression analysis: p values 0.09–0.72) at Week 2. Demographic variables — disease duration, age of onset — or variables known as severity markers — hip involvement, history of iritis — were not statistically related to the amount of change expressed by patients (regression analysis: p values 0.07–0.68).

ROC curve analyses were determined separately for each domain and for each change: absolute difference (Figure 1) or relative difference (Figure 2). Thus, in terms of absolute value, it was determined that a change superior or equal to 7 mm of BASFI corresponds to the smallest change, allowing the identification of a patient who experienced a clinically important improvement (sensitivity = 0.60; specificity = 0.85). For BASDAI the MCID value is 10 mm, with a sensitivity equal to 0.65 and specificity equal to 0.82. Finally, for BAS-G, this threshold value is 15 mm, with a sensitivity equal to 0.61 and specificity equal to 0.74 (Table 3).

DISCUSSION

The establishment of outcome criteria is of paramount importance when defining the natural history or response to treatment for any chronic disease. Thus, in terms of AS, we have disease activity, functional, and global definition (BASDAI, BASFI, BAS-G, respectively).

However, the next step relates to the translation of numerical values to readily understood concepts from both the patient's and clinician's perspective. Further, what has been shown to be statistically relevant must be shown to have clinical relevance. In this study, the minimum changes in BASDAI, BASFI, and BAS-G (MCID) that patients need to achieve to experience a meaningful improvement (rating of "good deal better" or more on a global rating scale) were determined.

This method allows for the use of these values to evaluate the effect of a therapeutic intervention at the patient level. In a clinical trial, groups can be compared by determining the number of patients who have experienced a "clinically important" improvement within every group. In essence, in contrast to many diseases for which laboratory tests are relevant, AS can only be readily defined by patient-related outcomes. These are typically self-administered instruments. Clearly, it is the patient who must play a central role in determining the value (or otherwise) of a new treatment modality.

Thus, the absence of a gold standard for assessing change in AS symptoms justifies the use of a global impression scale with a variety of terms defining improvement or deterioration of the patient's condition. Two similar global rating scales exist: one using 7 and the other using 15 points. The 15 point global rating scale, validated by Jaeschke, *et al*[6], was chosen because it is used most frequently in the literature. On this scale, the threshold value (or rather the threshold term defining a "clinically important" difference) is not unequivocal. Jaeschke, *et al* and Juniper, *et al*[15] defined the

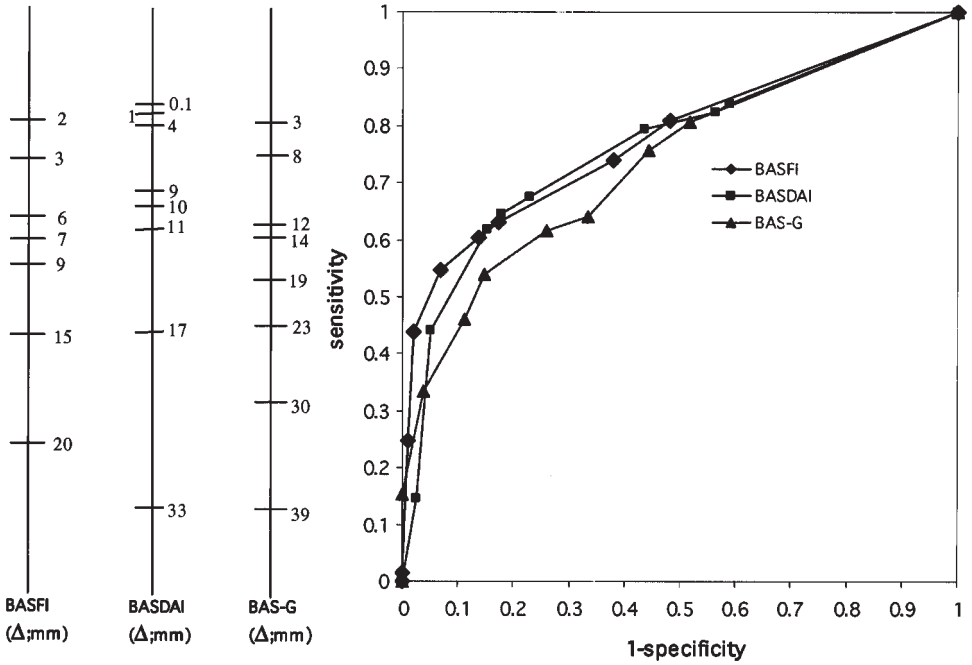

Figure 1. Receiver operating characteristic curves for the Bath Ankylosing Spondylitis indices based on absolute change.

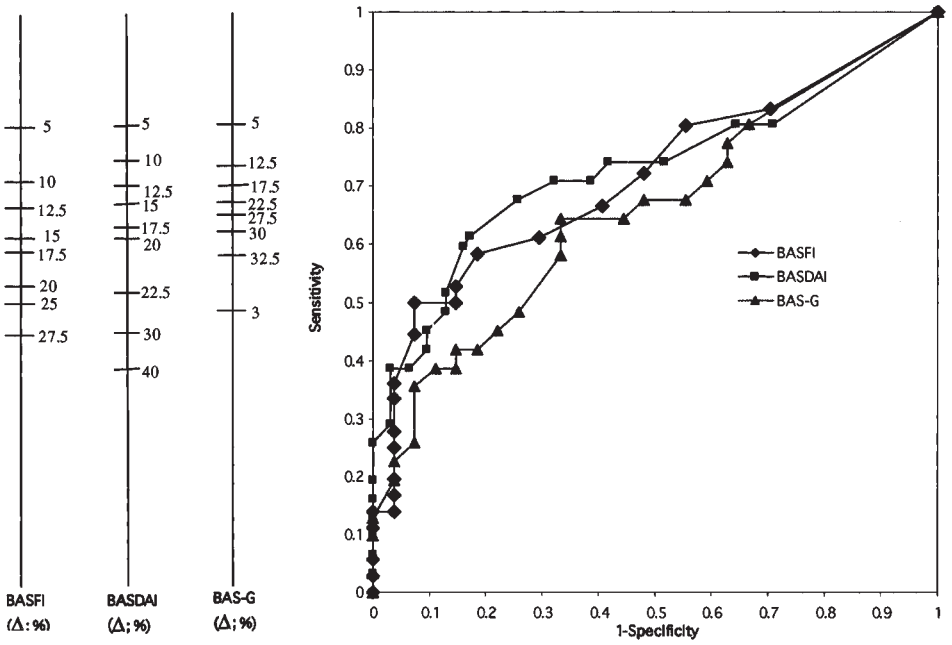

Figure 2. Receiver operating characteristic curves for the Bath Ankylosing Spondylitis indices based on relative change.

terms "little better" and "somewhat better" as threshold levels to determine the smallest clinically relevant difference. Stratford, *et al*[9] used the same global rating scale to determine the MCID of the Roland-Morris back pain questionnaire. Patients were regarded as having reached a clinically "important" improvement if they had classified their change in disability as "a good deal better" or "a great deal better" or "a very great deal better" following physiotherapy. The same choice as Stratford, *et al* was made because patients in this program were exposed to similar therapeutic interventions and we believe that a "clinically important" improvement must necessarily be classified in terms more

Table 3. Minimum clinically important difference values for AS outcome measures.

Outcome Measures	Minimum Clinically Important Difference	
	Absolute Value, mm	Relative Value, %
BASFI	7	17.50
	Sensitivity = 0.60, specificity = 0.85	Sensitivity = 0.60, specificity = 0.84
BASDAI	10	20.00
	Sensitivity = 0.65, specificity = 0.82	Sensitivity = 0.63, specificity = 0.78
BAS-G	15	27.50
	Sensitivity = 0.61, specificity = 0.74	Sensitivity = 0.64, specificity = 0.71

robust than "moderate." The choice of the threshold term underlines the difficulty in defining the concept of "clinically important" difference[16]. Does the smallest clinically perceptible difference represent a "clinically important" difference? For the practitioner, a "clinically important" difference can be perceived as the smallest change that would lead to a therapeutic strategy modification; however, when considering a treatment modification, the practitioner must consider the safety and cost of treatment. Thus, an improvement, classified as "moderate," might not be sufficiently robust to be considered "clinically important." Norman, *et al*[17] questioned the validity of a global rating scale. These authors have expressed doubts regarding the use of a global rating scale of unknown reliability composed of a single term to estimate specific tools consisting of multiple domains (e.g., function or disease activity). Previous work demonstrated that a specific measure, in a disease, is more sensitive to change than a generic measure[18]. Thus, one can think that a global rating scale does not estimate the same domain as the specific tool but reflects a more general status. However, it was observed in this study that patients separately judged the modifications of their health status according to the 3 domains studied. A second limitation, raised by Norman, *et al*, highlights the dependence of the global rating scale on patient recall of baseline disease state. Retrospectively, patients would frequently tend to exaggerate their baseline disease state. For example, Linton and Melin showed in a study on chronic pain that most patients rated the pain they had experienced previously as worse than they had initially noted[19]. The third limitation underlined by Norman, *et al* is that the measurement error of the global rating scale and the specific measure are correlated because both measures are simultaneously submitted to the patient.

In terms of absolute value, the MCID that were determined for BASDAI and BASFI are equivalent to the observed differences between groups in nonsteroidal antiinflammatory drug (NSAID) placebo-controlled trials with AS patients[20]. In this study, sensitivities and specificities obtained and the area under the ROC curves could be considered small. A sensitivity of 60% for MCID-BASFI may appear insufficient, and thus generate an error of classification. This low specificity and sensitivity is considered to be the result of the therapeutic intervention (e.g., physiotherapy) that the patients received. Although an important part of the treatment of AS, physiotherapy improves pain and stiffness, but has no effect on structural damage. Without modifying the obtained MCID values, it is hypothesized that the values of sensitivity and specificity would have been higher if the therapeutic intervention had been a drug (e.g., a NSAID).

No statistically significant association was found between initial scores and scores after 2 weeks. The use of the absolute MCID values is recommended rather than relative values, since the significance of a floor and ceiling effect will be more important for the latter. In this study, few patients experienced a worsening of their AS. Therefore, it was not possible to determine the MCID in terms of AS deterioration.

In summary, the values of the MCID for 3 outcomes measures commonly used in the spondyloarthropathies were defined: 10 mm for BASDAI (disease activity), 7 mm for BASFI (function), and 15 mm for BAS-G (global well being). These values make the evaluation of AS more readily understandable for both patients and clinicians, and relate to the evaluation of therapy at the patient level. More research is necessary to validate these MCID values among similar groups of patients with AS. Further, the development of very effective therapies for patients with AS, such as antitumor necrosis factor drugs, will require establishing definitions of partial and total remission of AS.

REFERENCES

1. Ruof J, Stucki G. Validity aspects of erythrocyte sedimentation rate and C-reactive protein in ankylosing spondylitis. J Rheumatol 1999;26:966-70.
2. Garrett S, Jenkinson T, Kennedy LG, Whitelock H, Gaisford P, Calin A. A new approach to defining disease status in ankylosing spondylitis: The Bath Ankylosing Spondylitis Disease Index. J Rheumatol 1994;21:2286-91.
3. Calin A, Garrett SL, Whitelock H, et al. A new approach to defining functional ability in ankylosing spondylitis: The development of the Bath Ankylosing Spondylitis Function Index. J Rheumatol 1994;21:2281-5.
4. Jones S, Steiner A, Garret SL, Calin A. The Bath Ankylosing Spondylitis patient Global score (BAS-G). Br J Rheumatol 1996;35:66-71.
5. Van der Heijde D, van der Linden S, Dougados M, Bellamy N, Russel A, Edmonds J. Ankylosing spondylitis: Plenary discussion and voting on selection of domains and some specific instruments. J Rheumatol 1999;26:1003-5.

6. Jaeschke R, Singer J, Guyatt G. Measurement of health status. Ascertaining the minimal clinically important difference. Control Clin Trials 1989;10:407-15.
7. Wells G, Beaton D, Shea B. Minimal clinically important differences: Review of methods. J Rheumatol 2001;28:406-12.
8. Wells G, Anderson J, Beaton D. Minimal clinically important differences module: Summary, recommendations, and research agenda. J Rheumatol 2001;28:452-5.
9. Stratford PW, Binkley JM, Riddle DL, Guyatt GH. Sensitivity to change of the Roland-Morris Back Pain Questionnaire. Phys Ther 1998;11:1186-207.
10. Felson DT, Anderson JJ. A review of evidence on the discriminant validity of outcome measures in rheumatoid arthritis. J Rheumatol 2001;28:422-6.
11. Cranney A, Welch V, Wells G, et al. Discrimination of changes in osteoporosis outcomes. J Rheumatol 2001;28:413-21.
12. Ravaud P, Giraudeau B, Audeley R, Edouard-Noel R, Dougados M, Chastang C. Assessing smallest detectable change over time in continuous structural outcome measures: Application to radiological change in knee osteoarthritis. J Clin Epidemiol 1999;52:1225-30.
13. Van der Linden S, Valkenburg H, Cats A. Evaluation of diagnostic criteria for ankylosing spondylitis, a proposal for modification of the New York criteria. Arthritis Rheum 1984;27:361-8.
14. Wolfe F. Comparative usefulness of C-reactive protein and erythrocyte sedimentation rate in patients with RA. J Rheumatol 1997;24:1477-85.
15. Juniper E, Guyatt G, Willan A, Griffith L. Determining a minimal important change in a disease-specific quality of life questionnaire. J Clin Epidemiol 1994;47:81-7.
16. Beaton DE, Boers M, Wells GA. Many faces of the minimal clinically important difference (MCID): a literature review and directions for future research. Curr Opin Rheumatol 2002;14:109-14.
17. Norman GR, Stratford PW, Regehr G. Methodological problems in the retrospective computation of responsiveness to change: the lesson of Cronbach. J Clin Epidemiol 1997;50:869-79.
18. Tugwell P, Bombardier C, Buchanan W, Goldsmith C, Grace E, Hanna B. The MACTAR Patient Preference Disability Questionnaire: an individualized functional priority approach for assessing improvement in physical disability in clinical trials in rheumatoid arthritis. J Rheumatol 1987;14:446-51.
19. Linton SJ, Melin L. The accuracy of remembering chronic pain. Pain 1982;13:281-5.
20. Dougados M, Béhier JM, Jolchine I, et al. Efficacy of celecoxib in the treatment of ankylosing spondylitis. Arthritis Rheum 2001;44:180-5.