

## EXTENDED REPORT

# Evaluation of clinically relevant changes in patient reported outcomes in knee and hip osteoarthritis: the minimal clinically important improvement

F Tubach, P Ravaud, G Baron, B Falissard, I Logeart, N Bellamy, C Bombardier, D Felson, M Hochberg, D van der Heijde, M Dougados



*Ann Rheum Dis* 2005;64:29–33. doi: 10.1136/ard.2004.022905

See end of article for authors' affiliations

Correspondence to:  
Dr F Tubach, Département d'Epidémiologie, Biostatistique et Recherche Clinique, INSERM E0357, Hôpital Bichat, 46 rue Henri Huchard, 75018 Paris, France; florence.tubach@bch.ap-hop-paris.fr

Accepted 9 May 2004  
Published Online First 18 June 2004

**Background:** In clinical trials, at the group level, results are usually reported as mean and standard deviation of the change in score, which is not meaningful for most readers.

**Objective:** To determine the minimal clinically important improvement (MCII) of pain, patient's global assessment of disease activity, and functional impairment in patients with knee and hip osteoarthritis (OA).

**Methods:** A prospective multicentre 4 week cohort study involving 1362 outpatients with knee or hip OA was carried out. Data on assessment of pain and patient's global assessment, measured on visual analogue scales, and functional impairment, measured on the Western Ontario McMaster Universities Osteoarthritis Index (WOMAC) function subscale, were collected at baseline and final visits. Patients assessed their response to treatment on a five point Likert scale at the final visit. An anchoring method based on the patient's opinion was used. The MCII was estimated in a subgroup of 814 patients (603 with knee OA, 211 with hip OA).

**Results:** For knee and hip OA, MCII for absolute (and relative) changes were, respectively, (a) –19.9 mm (–40.8%) and –15.3 mm (–32.0%) for pain; (b) –18.3 mm (–39.0%) and –15.2 mm (–32.6%) for patient's global assessment; (c) –9.1 (–26.0%) and –7.9 (–21.1%) for WOMAC function subscale score. The MCII is affected by the initial degree of severity of the symptoms but not by age, disease duration, or sex.

**Conclusion:** Using criteria such as MCII in clinical trials would provide meaningful information which would help in interpreting the results by expressing them as a proportion of improved patients.

The choice of an outcome measure is a major step in the design of clinical trials. In evaluating the symptomatic severity of osteoarthritis (OA) of the lower limbs, scientific groups such as the OMERACT (Outcome Measures in Rheumatology Group),<sup>1</sup> GREES (Group for the Respect of Ethics and Excellence in Science),<sup>2</sup> and OARSI (Osteoarthritis Research Society International)<sup>3</sup> have raised the importance of evaluating at least three dimensions: pain, patient's global assessment of disease status, and functional impairment. At the individual level, determining the minimal meaningful change in a score by use of a structured instrument is a challenge. Are changes in self reported levels of pain of 10 mm on a 0–100 mm visual analogue scale (VAS) clinically important? Does the change reflect meaningful improvement for the patient? The concept of the minimal clinically important difference (MCID)<sup>4–6</sup> could help in interpreting changes in scores at the individual level. However, the MCID, which can reflect either an improvement or a worsening, has not been used here, because in clinical trials we are always interested in improvement and not worsening. Furthermore, it has been shown that the MCID could be different for improvement and worsening.<sup>7</sup> The minimal clinically important improvement (MCII), defined as the smallest change in measurement that signifies an important improvement in a patient's symptom, seems more appropriate and, in clinical trials, provides readers with additional information on the effect size by expressing the results more meaningfully (that is, as a percentage of improved patients).

This prospective cohort study aimed at estimating the MCII from the patient's perspective for three main patient reported

outcomes used in OA trials: pain, patient's global assessment of disease activity, and functional impairment.

## MATERIALS AND METHODS

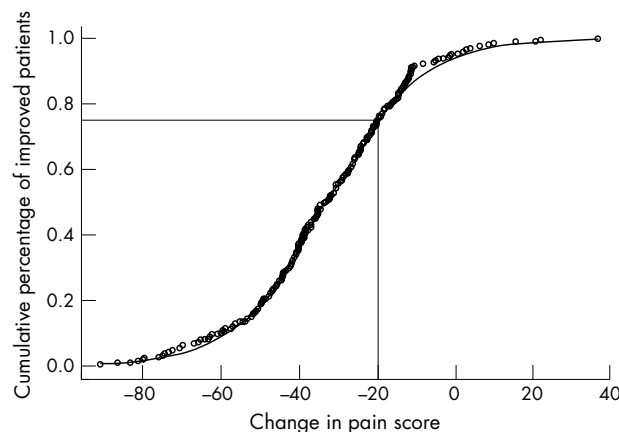
### Study design

We conducted a prospective 4 week cohort study.

### Study population

This study involved 1362 outpatients with knee or hip OA, as defined by the American College of Rheumatology,<sup>8,9</sup> included by 399 rheumatologists. Each rheumatologist had to recruit four patients, three with knee OA and one with hip OA. To be included in the study, patients had to experience pain from OA ( $\geq 30$  mm on a VAS varying from 0 to 100), require treatment with a non-steroidal anti-inflammatory drug (NSAID), and be able to complete questionnaires in French. Inclusion could begin with the onset of treatment or a switch from one NSAID to another. Patients were excluded if they had a prosthesis on the assessed joint or if they had been given an intra-articular injection in the 4 weeks before the study began. All patients initially visited the rheumatologist in charge of their case, and an NSAID was prescribed (the drug and its dosage was chosen by the physician). A final visit to the same rheumatologist was scheduled 4 weeks later.

**Abbreviations:** MCID, minimal clinically important difference; MCII, minimal clinically important improvement; NSAID, non-steroidal anti-inflammatory drug; OA, osteoarthritis; VAS, visual analogue scale; WOMAC, Western Ontario McMaster Universities Osteoarthritis Index



**Figure 1** Aspects of the cumulative distribution function used to determine the MCII (changes in pain score in patients with knee OA;  $n = 265$ ). Among patients considering their response to treatment as good on a five point Likert scale, 75% experienced a decrease in pain between baseline and final visit of  $>19.9$  mm on a 0–100 mm VAS (a change between  $-100$  mm and  $-19.9$  mm).

### Measurements

At the baseline visit, demographic and disease data were collected. Patients assessed their OA status at baseline and final visit. They assessed the following patient reported outcomes: (a) pain on movement during the 48 hours before the visit, measured on a 0–100 mm VAS; (b) global assessment of disease activity measured on a 0–100 mm VAS; and (c) physical function, measured on the Western Ontario McMaster Universities Osteoarthritis Index (WOMAC) function subscale (17 items, five point Likert scale for each item; high scores indicate high degree of functional impairment; total score normalised to a 0–100 score).

At the final visit, a random sample of two thirds of the patients ( $n = 923$ ) assessed their response to NSAID

treatment on a five point Likert scale (none = no good at all, ineffective drug; poor = some effect but unsatisfactory; fair = reasonable effect but could be better; good = satisfactory effect with occasional episodes of pain or stiffness; excellent = ideal response, virtually pain free). The other third of the patients assessed their response to treatment on a 15 point Likert scale (from  $-7$ , a very great deal worse, to  $+7$ , a very great deal better, with 0, no change).

### Statistical analysis

All the analyses considered patients with knee and hip OA separately.

The MCII was determined in a subgroup of 814 patients (603 with knee and 211 with hip OA) whose assessment of response to treatment was measured on a five point Likert scale and who had completed the final visit.

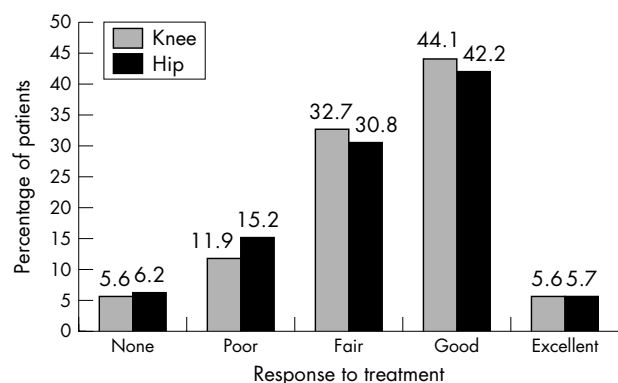
An anchoring method based on the patient's assessment of response to treatment was used.

The MCII was estimated for both the absolute (final value–baseline value) and the relative ((final value–baseline value)/baseline value) changes in each patient reported outcome. It was estimated by constructing a curve of cumulative percentages of patients as a function of the change in score (for example, difference in pain score) among patients whose final evaluation of response to treatment was “good, satisfactory effect with occasional episodes of pain or stiffness”, because we wanted to focus on the improvement that was clinically important. Logistic regression was used to model the observations (fig 1). We targeted the point at the flattening of the curve at which most subjects stated they had improved. To determine the change in score corresponding to this point, we first looked at the two parameter logistic model that best fitted the data. Then we determined the square root of the third derivative of this logistic function that corresponded with the MCII. One can demonstrate that this point corresponds by construction to the 78.9th centile of the change in score, and thus we propose to define the MCII as the 75th centile of the change in score, because it is very close to the point defined above and easier to derive. The model permitted us, firstly, to determine that the target point was

**Table 1** Baseline characteristics of patients

Characteristics	Knee OA (n=603)		Hip OA (n=211)	
	Mean	SD	Mean	SD
Age (years)	67.9	10.2	64.6	10.2
Weight (kg)	75.5	13.8	71.3	12.3
Height (cm)	163.8	8.5	164.7	8.4
Body mass index (kg/m <sup>2</sup> )	28.1	4.8	26.2	3.8
Disease duration (years)	4.7	5.8	3.3	4.6
Pain score (0–100 mm VAS)				
Week 0	59.3	16.2	56.7	16.5
Change (week 0–week 4)	–24.9	21.5	–20.0	21.7
Patient global assessment (0–100 mm VAS)				
Week 0	59.6	18.3	58.0	19.3
Change (week 0–week 4)	–24.7	24.0	–20.6	23.2
WOMAC function score (0–100)				
Week 0	42.8	16.1	44.4	16.5
Change (week 0–week 4)	–11.6	13.9	–10.4	13.6
Female sex	No	%	No	%
Kellgren & Lawrence grade				
II	421	69.8	133	63.0
III	108	17.9	33	15.7
IV	268	44.4	111	52.9
IV	227	37.7	66	31.4
NSAID* intake during past 4 weeks	178	29.7	69	32.7
Analgesic treatment**	344	57.2	14.1	67.1
Symptomatic slow-acting drug intake***	209	34.8	90	42.9

\*Non-steroidal anti-inflammatory drugs (before the start of the study); †other than NSAIDs (before the start of the study); ‡chondroitin sulphate, diacerhein, or avocado/soybean unsaponifiables.



**Figure 2** Patients' assessment of their response to treatment.

correctly approached by the 75th centile and, secondly, to estimate the 95% confidence intervals.

In a second step, we stratified the analysis on the baseline score of interest (divided into tertiles) to assess whether the level of pain, the patient's assessment of disease activity, and functional impairment had a modifying effect on the MCII. That is we stratified (*a*) on the baseline pain score to estimate the MCII for pain; (*b*) on the baseline assessment of disease activity to estimate the MCII for patient's assessment of disease activity; (*c*) on the baseline WOMAC function score to estimate the MCII for functional impairment.

In a third step, to investigate the effect of covariates (other than location of OA) on the MCII, we stratified the analysis successively by age, disease duration (both divided into tertiles), and sex.

Statistical analyses was performed with the SAS Release 8.2 statistical software package and the S plus 4.5 statistical software package.

### Compliance with research ethics standards

This study was conducted in compliance, with the protocol, good clinical practices, and the Declaration of Helsinki principles.

### RESULTS

A total of 1362 patients were enrolled in the study: 1019 (75%) had knee and 343 (25%) hip OA; 913 (67%) were female; and the mean (SD) age was 67.2 (10.5) years. A total of 914 (90%) patients with knee and 310 (90%) with hip OA completed the final visit. Patients lost to follow up were excluded from the analysis and did not differ from completers in their baseline characteristics. Among the completers, 603 patients with knee and 211 with hip OA assessed their response to treatment on a five point Likert scale.

Table 1 shows the descriptive statistics on clinical and demographics variables. Figure 2 shows patients' rating of response to treatment.

Table 2 lists the MCII values for the three patient reported outcomes, according to location of OA. These values were estimated in the 265 patients with knee and the 87 patients with hip OA who completed the final visit and assessed their response to treatment as "good". For instance, patients with knee OA considered themselves clinically improved if the decrease in pain exceeded 19.9 mm on the 0–100 mm VAS. We used the data from the five point not the 15 point Likert scale mentioned in the "Methods" section.

Table 3 shows the estimates of the MCII (for absolute change) stratified on the baseline score in patients with knee or hip OA. The higher the baseline score, the larger the MCII. Patients who have a severe symptom need a higher level of change to consider themselves clinically improved than those with less severe symptoms. For instance, patients with severe pain (a high tertile of baseline pain score) considered themselves clinically improved if the decrease in pain exceeded 36.6 mm on the 0–100 mm VAS. Patients with less pain (low tertile of baseline pain score) needed a lower level of change (–10.8 mm on the VAS) to consider themselves clinically improved. The estimates of the MCII for relative change also varied across tertiles of the baseline score (data not shown).

The estimates of the MCII do not vary across age, disease duration tertiles, or sex (data not shown).

### DISCUSSION

This study dealt with the clinical meaningfulness of changes observed for patient reported outcome measures. Because a statistically significant difference is mostly a matter of sample size, the most difficult issue is whether an observed or estimated difference is clinically important.<sup>10</sup> In other words, statistical significance is not equivalent to clinical significance. Reporting results of a trial using the MCII (that is, as a percentage of improved patients) provides readers with values which are more easily understood and additional information to help them decide whether a treatment should be used. This threshold also allows for monitoring of individual response to treatment over time and adapting treatment to individual patients (for example, determining whether to start or interrupt a treatment). Furthermore, the designation and use of MCII in clinical trials is critical for meaningful systematic reviews and combining results from different studies in meta-analyses.<sup>11</sup> This concept aims at complementing, not replacing, information on the effect size, because the effect size remains a more powerful approach.<sup>12</sup>

The MCII is the smallest change in measures that signifies an important improvement in a patient's symptom. Thus, the MCII can undoubtedly be considered as a treatment target from the patient's perspective. It is based on the patient's opinion as an external anchor and contrasts changes within

**Table 2** Minimal clinically important improvement (MCII) scores according to patients' location of OA

Patient reported outcomes	Knee OA				Hip OA			
	Absolute change		Relative change		Absolute change		Relative change	
	MCII	(95% CI)	MCII (%)	(95% CI)	MCII	(95% CI)	MCII (%)	(95% CI)
Pain (0–100 mm VAS), mm	–19.9	(–21.6 to –17.9)	–40.8	(–44.8 to –36.1)	–15.3	(–17.8 to –12.5)	–32.0	(–38.5 to –24.0)
Patient global assessment (0–100 mm VAS), mm	–18.3	(–19.8 to –16.7)	–39.0	(–45.8 to –30.6)	–15.2	(–16.9 to –13.4)	–32.6	(–38.7 to –25.2)
WOMAC function score (0–100)	–9.1	(–10.5 to –7.5)	–26.0	(–28.6 to –23.3)	–7.9	(–8.8 to –5.0)	–21.1	(–24.8 to –17.0)

The MCII was defined as the 75th centile of the change in score among patients whose evaluation of response to treatment was "good", for the three patient reported outcomes: pain, as assessed on a visual analogue scale (VAS); global assessment of disease status, on a VAS; or the Western Ontario McMaster Universities Osteoarthritis Index (WOMAC) function subscale score.

**Table 3** Minimal clinically important improvement (MCII) score of absolute change in patients with knee or hip OA, by low, intermediate, and high baseline score tertiles

	Knee OA			Hip OA		
	Baseline score tertile			Baseline score tertile		
	Low	Intermediate	High	Low	Intermediate	High
Pain (0–100 mm VAS)	–10.8 (–12.7 to –8.7) {30 to 51.0}	–27.4 (–29.7 to –24.6) {51.1 to 66.2}	–36.6 (–38.3 to –34.7) {>66.2}	–7.2 (–10.7 to –2.9) {30 to 49.3}	–23.9 (–28.3 to –18.0) {49.4 to 65.4}	–29.7 (–35.4 to –21.8) {>65.4}
Patient's global assessment of disease (0–100 mm VAS)	–6.4 (–8.6 to –3.8) {≤50.4}	–24.6 (–26.8 to –22.1) {50.5 to 68.7}	–43.2 (–46.8 to –38.7) {>68.7}	–4.3 (–6.9 to –1.4) {≤49.9}	–26.0 (–28.3 to –23.3) {50.0 to 69.9}	–29.9 (–34.5 to –24.3) {>69.9}
WOMAC function score (0–100)	–5.3 (–6.5 to –3.8) {≤35.3}	–11.8 (–13.0 to –10.4) {35.4 to 51.5}	–20.4 (–22.5 to –18.1) {>51.5}	–2.6 (–4.4 to –0.5) {≤38.2}	–14.8 (–17.0 to –12.0) {38.3 to 52.9}	–15.1 (–18.9 to –10.0) {>52.9}

Results are shown as mean (95% confidence interval) {tertile}.

The MCII was defined as the 75th centile of the change in score among patients whose evaluation of response to treatment was “good”, for three patient reported outcomes: pain, as assessed on a visual analogue scale (VAS); global assessment of disease status, on a VAS; or the Western Ontario McMaster Universities Osteoarthritis Index (WOMAC) function subscale score.

patients at the individual level (proportion of improved patients) instead of at the group level (mean change in a variable).

Approaches such as investigator defined (expert consensus) or statistically defined methods have been used to determine this threshold.<sup>13,14</sup> Despite the absence of a criterion measure, establishing the meaning of changes in a measure requires an independent standard. Patient global ratings are recommended as an external anchor for evaluating the clinical significance of individual change.<sup>13</sup> The large sample of patients as experts in determining improvement is a good indicator of representativeness.

To determine the MCII, the external criterion was the patient's assessment of response to treatment as assessed on a five point Likert scale. We defined MCII in the group of patients whose evaluation of response to treatment was “good”, because one is always looking for clinically important differences. We did not include patients whose evaluation of response to treatment was “excellent,” because our target was the *minimal* change important from the patient's perspective. But obviously, this choice was arbitrary and affects the results (data not shown). The group of patients in whom MCII is determined and the wording of the items in the questionnaire to assess response to treatment should be chosen with the help of experts; in our study, the group of patients were chosen by the experts NB, CB, DF, MH, DvdH, and MD.

In a previous study,<sup>15</sup> a three round Delphi method involving six academic rheumatologists experienced in OA trials was used to define the MCID for some outcome measures used in OA trials (not specifically focusing on hip or knee OA). The MCID for patient pain on movement (measured on a 0–100 VAS) was 17.5 mm and that for patient global assessment of disease activity (measured on a 0–100 VAS) was 15 mm. Although this method differs from that used in our study, the values are very close to our estimates of MCII for these patient reported outcomes. The only study dealing with meaningful change for the WOMAC dealt with the minimal clinically perceptible difference not the MCID.<sup>16</sup>

Our study has demonstrated that the MCII varies depending on the baseline state. Patients who have the most severe symptoms have to experience a greater change to consider themselves improved. Riddle and colleagues also found this effect in their investigation of low back pain,<sup>17</sup> where the MCID varied between 3 and 13 depending on the baseline range of scores (on the Roland-Morris Back Pain Questionnaire,<sup>18</sup> total score varying from 0 to 24 points, with

baseline scores divided into five approximately equal sized intervals). However, the precision of their estimates may have been compromised by the small sample size, especially for patients with high levels of disability.

The variation of MCII across tertiles of baseline scores in our study cannot be imputed to the size of the sample, as confirmed by the narrowness of the 95% confidence intervals. We believe that this variation depending on the baseline score may preclude the use of the crude MCII. The patient's initial or previous score should be taken into account when making decisions about important change. We propose to use three estimates of MCII, corresponding to the tertiles of each baseline score, to express the changes in terms of important improvement. This meets the recommendation of Crosby and associates<sup>13</sup> for estimating MCID in health related quality of life criteria: to anchor baseline severity of individual patients.

We believe this is the first study to investigate the effect of several covariates such as age, sex, OA location, and disease duration on patient responses. It is interesting to observe that these factors do not consistently modify the estimates of MCII.

In conclusion, use of the concept MCII facilitates the presentation and interpretation of results obtained in clinical trials and the transposition of trial results into practice. However, the baseline score should be taken into account. Further studies involving different datasets, clinical environments, languages, and countries are necessary to validate these observations prospectively.

## ACKNOWLEDGEMENTS

This study was supported by an unrestricted grant from Merck, Sharp, & Dohme Chibret Laboratories, France.

## Authors' affiliations

**F Tubach, P Ravaud, G Baron**, Institut National de la Santé et de la Recherche Médicale (INSERM) E 0357; Département d'Epidémiologie, Biostatistique et Recherche Clinique; Groupe Hospitalier Bichat-Claude Bernard (Assistance Publique—Hôpitaux de Paris); Faculté Xavier Bichat (Université Paris 7), Paris, France

**B Falissard**, Faculté de Médecine Paris-sud, Département de Santé Publique, Hôpital Paul Brousse (Assistance Publique—Hôpitaux de Paris), Villejuif, France

**I Logeart**, Merck, Sharp & Dohme Chibret Laboratories, Paris, France

**N Bellamy**, Department of Medicine, University of Queensland, Royal Brisbane Hospital, Brisbane, Queensland, Australia

**C Bombardier**, Institute for Work and Health, Toronto, Ontario, Canada

**D Felson**, Boston University School of Medicine, Boston, Massachusetts, USA

**M Hochberg**, University of Maryland, Baltimore, Maryland, USA  
**D van der Heijde**, University Hospital, Maastricht, The Netherlands  
**M Dougados**, Service de Rhumatologie B, Hôpital Cochin (Assistance Publique—Hôpitaux de Paris), Paris, France

## REFERENCES

- Bellamy N**, Kirwan J, Boers M, Brooks P, Strand V, Tugwell P, *et al*. Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip, and hand osteoarthritis. Consensus development at OMERACT III. *J Rheumatol* 1997;**24**:799–802.
- GREES**. Recommendations for the registration of drugs used in the treatment of osteoarthritis. Group for the respect of ethics and excellence in science (GREES): osteoarthritis section, *Ann Rheum Dis* 1996;**55**:552–7.
- Altman R**, Brandt K, Hochberg M, Moskowitz R, Bellamy N, Bloch DA, *et al*. Design and conduct of clinical trials in patients with osteoarthritis: recommendations from a task force of the Osteoarthritis Research Society. Results from a workshop. *Osteoarthritis Cartilage* 1996;**4**:217–43.
- Beaton DE**, Bombardier C, Katz JN, Wright JG. A taxonomy for responsiveness. *J Clin Epidemiol* 2001;**54**:1204–17.
- Juniper EF**, Guyatt GH, Willan A, Griffith LE. Determining a minimal important change in a disease-specific Quality of Life Questionnaire. *J Clin Epidemiol* 1994;**47**:81–7.
- Jaeschke R**, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials* 1989;**10**:407–15.
- Schwartz AL**, Meek PM, Nail LM, Fargo J, Lundquist M, Donofrio M, *et al*. Measurement of fatigue. determining minimally important clinical differences. *J Clin Epidemiol* 2002;**55**:239–44.
- Altman R**, Alarcon G, Appelrouth D, Bloch D, Borenstein D, Brandt K, *et al*. The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hip. *Arthritis Rheum* 1991;**34**:505–14.
- Altman R**, Asch E, Bloch D, Bole G, Borenstein D, Brandt K, *et al*. Development of criteria for the classification and reporting of osteoarthritis. Classification of osteoarthritis of the knee. Diagnostic and Therapeutic Criteria Committee of the American Rheumatism Association. *Arthritis Rheum* 1986;**29**:1039–49.
- Wright JG**. Interpreting health-related quality of life scores: the simple rule of seven may not be so simple. *Med Care* 2003;**41**:597–8.
- Brooks P**, Tugwell P, Strand CV, Simon L. International Consensus Conference on Outcome Measures in Rheumatology, Gold Coast, Australia, April 11–14, 2002. *J Rheumatol* 2003;**30**:866–7.
- Streiner DL**. Breaking up is hard to do: the heartbreak of dichotomizing continuous data. *Can J Psychiatry* 2002;**47**:262–6.
- Crosby RD**, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. *J Clin Epidemiol* 2003;**56**:395–407.
- Beaton DE**. Simple as possible? Or too simple? Possible limits to the universality of the one half standard deviation. *Med Care* 2003;**41**:593–6.
- Bellamy N**, Carrette S, Ford PM, Kean WF, le Riche NG, Lussier A, *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.
- Ehrich EW**, Davies GM, Watson DJ, Bolognese JA, Seidenberg BC, Bellamy N. Minimal perceptible clinical improvement with the Western Ontario and McMaster Universities osteoarthritis index questionnaire and global assessments in patients with osteoarthritis. *J Rheumatol* 2000;**27**:2635–41.
- Riddle DL**, Stratford PW, Binkley JM. Sensitivity to change of the Roland-Morris Back Pain Questionnaire: part 2. *Phys Ther* 1998;**78**:1197–207.
- Roland M**, Morris R. A study of the natural history of back pain. Part I: development of a reliable and sensitive measure of disability in low-back pain. *Spine* 1983;**8**:141–4.