

## EXTENDED REPORT

# Evaluation of WOMAC 20, 50, 70 response criteria in patients treated with hylan G-F 20 for knee osteoarthritis

N Bellamy, M J Bell, C H Goldsmith, D Pericak, V Walker, J P Raynauld, G W Torrance, P Tugwell, R Polisson



*Ann Rheum Dis* 2005;64:881–885. doi: 10.1136/ard.2004.026443

See end of article for authors' affiliations

Correspondence to: Professor Nicholas Bellamy, CONROD, Faculty of Health Sciences, University of Queensland, Level 3, Mayne Medical School, Herston Road, Brisbane, Queensland 4006, Australia; nbellamy@medicine.uq.edu.au

Accepted 27 October 2004  
Published Online First  
25 November 2004

**Objective:** A secondary analysis of a previously conducted one year randomised controlled trial to evaluate the capacity of responder criteria based on the WOMAC index to detect between treatment group differences.

**Methods:** 255 patients with knee osteoarthritis were randomised to "appropriate care with hylan G-F 20" (AC+H) or "appropriate care without hylan G-F 20" (AC). In the original analysis, two definitions of patient response from baseline to month 12 were used: (1) at least a 20% reduction in WOMAC pain score (WOMAC 20P); (2) at least a 20% reduction in WOMAC pain score and at least a 20% reduction in either WOMAC function or stiffness score (WOMAC 20PFS). For this analysis, a responder was identified using 50% and 70% minimum clinically important response levels to investigate how increasing response affects the ability to detect treatment group differences.

**Results:** The hylan G-F 20 group had numerically more responders using all patient responder criteria. Increasing the response level from 20% to 50% detected similar differences between treatment groups (25% to 29%). Increasing the response level to 70% reduced the differences between treatment groups (11% to 12%) to a point where the differences were not significant after Bonferroni adjustment.

**Conclusions:** These results provide evidence for incorporating response levels (WOMAC 50) in clinical trials. While differences at the highest threshold (WOMAC 70) were not statistically detectable, an appropriately powered study may be capable of detecting differences even at this very high level of improvement.

Developments in standardisation of outcome measurement procedures for clinical trials in the treatment of osteoarthritis and rheumatoid arthritis have followed similar but not identical pathways. While measures of pain, function, and patient global assessment have been selected as core set measures for clinical trials in both of these diseases, the outcome measures in arthritis clinical trials—American College of Rheumatology (OMERACT-ACR) criteria for rheumatoid arthritis<sup>1</sup> differ from the OMERACT-Osteoarthritis Research Society International (OARSI) criteria for osteoarthritis,<sup>2,3</sup> in that the former also include measures of the number of tender and swollen joints, physician global assessment, and C reactive protein/erythrocyte sedimentation rate values. The subsequent development of responder criteria for rheumatoid arthritis<sup>4</sup> and osteoarthritis trials<sup>5,6</sup> reflects these differences in core set measures. In addition, ACR responder criteria for rheumatoid arthritis<sup>4</sup> are based on percentage changes on two or more variables, while OMERACT-OARSI responder criteria for osteoarthritis<sup>6</sup> are based on a combination of percentage and absolute changes in one or more variables. Following the development of the ACR 20 responder criteria for rheumatoid arthritis,<sup>4</sup> higher threshold requirements for response designation have been explored, namely ACR 50 and ACR 70 responder criteria.<sup>7</sup> Higher response levels have been more difficult to achieve, and between group differences in rheumatoid arthritis clinical trials have (albeit less often) been detected at these higher thresholds, requiring individual patient improvements at or above the 50% and 70% levels, respectively.<sup>8–13</sup>

Notwithstanding the principle of employing a combination of percentage and absolute changes of one or more variables

in OMERACT-OARSI responder criteria for osteoarthritis,<sup>6</sup> and that of basing responder criteria on percentage change alone in OMERACT-ACR criteria for rheumatoid arthritis,<sup>4</sup> we undertook secondary analyses of a published randomised controlled trial<sup>14,15</sup> to evaluate the ability of responder criteria based on the Western Ontario and McMaster Universities (WOMAC) osteoarthritis index to detect between treatment differences. We compared the results of analyses based on WOMAC 20, WOMAC 50, and WOMAC 70 responder criteria to determine whether the application of different criteria influences data interpretation.

## METHODS

### Design

The analyses reported here were undertaken using the data collected in a health outcomes trial evaluating viscosupplementation with hylan G-F 20 when added to an appropriate care treatment regimen for patients with knee osteoarthritis. The detailed design of this trial and the primary analyses of the data have been published elsewhere.<sup>14,15</sup> Briefly, the trial was a multicentre, randomised, controlled, open label study over one year, where patients were randomised to either "appropriate care with hylan G-F 20" (AC+H) or to "appropriate care without hylan G-F 20" (AC). Appropriate care for knee osteoarthritis was defined by the guidelines for the medical management of osteoarthritis proposed by the ACR.<sup>16</sup> Patients in this study had symptomatic knee osteoarthritis (of mild to moderate severity) and had received previous treatment with analgesics. Appropriate care could

**Abbreviations:** ACR, American College of Rheumatology; WOMAC, Western Ontario and McMaster Universities osteoarthritis index

include treatment with analgesics, non-steroidal anti-inflammatory drugs (NSAIDs), corticosteroid injections, supportive measures such as education and counselling, weight loss, joint rest, application of heat or ice, use of devices, physical therapy, arthroscopy, and total joint replacement. Patients randomised to the AC+H group could receive more than one course of hylan G-F 20 treatment in the study knee (the knee that was most symptomatic or with the predominant musculoskeletal problem) if medically warranted, and could receive bilateral treatment if their contralateral knee was affected. Retreatment was provided when persistent pain recurred, with a minimum of four weeks between courses of hylan G-F 20. Patients were assessed by the clinical investigator at baseline and at 12 months. Follow up assessments were completed by telephone at months 1, 2, 4, 6, 8, 10, and 12. The study protocol and informed consent form were approved by the relevant ethics committees for the sites. Informed consent was obtained from each patient.

**Outcome measures**

The WOMAC Likert 3.0 is a self administered, disease specific health related quality of life instrument that asks the patients questions concerning the study knee. It produces one aggregate total score and scores for three subscales: pain, stiffness, and physical functioning. A higher score for each subscale corresponds to a worse condition. The pain subscale includes five questions on the degree of pain experienced with certain positions and activities (for example, sitting or lying), with the subscore varying from 0 to 20. The function subscale includes 17 questions on the degree of difficulty experienced while completing activities (for example, descending stairs); the subscore varies from 0 to 68. The stiffness subscale includes two questions on severity of stiffness (that is, after first awakening, and later in the day), with the subscore varying from 0 to 8. For every question in the WOMAC, patients rate their pain, stiffness, or function using five ordinal responses: none, mild, moderate, severe, and extreme. The WOMAC was completed in the office at baseline and by telephone at months 1, 2, 4, 6, 8, 10, and 12.

**Table 1** Demographic variables and disease status at baseline

| Baseline measure                                   | AC+H (n = 127) | AC (n = 128) |
|--|----------------|--------------|
| Age (years)  | 62.6 (9.4)     | 63.5 (10.5)  |
| Body mass index (kg/m <sup>2</sup> )               | 32.1 (8.0)     | 32.9 (7.2)   |
| Duration (years) of OA symptoms                    |                |              |
| Study knee   | 9.0 (9.5)      | 9.9 (9.7)    |
| Other knee   | 7.4 (8.8)      | 8.3 (9.3)    |
| Sex: female  | 86 (68%)       | 93 (73%)     |
| Previous treatment for OA of the knee(s)           |                |              |
| Acetaminophen                                      | 100 (79%)      | 109 (85%)    |
| NSAIDs   | 120 (94%)      | 110 (86%)    |
| Previous surgery, study knee                       | 40 (31%)       | 39 (30%)     |
| Previous surgery, other knee                       | 40 (31%)       | 23 (18%)     |
| Radiology grading within 1 year (central grading*) |                |              |
| Not reported                                       | 0 (0%)         | 1 (1%)       |
| Grade 0  | 4 (3%)         | 4 (3%)       |
| Grade I  | 17 (13%)       | 11 (9%)      |
| Grade II   | 32 (25%)       | 33 (26%)     |
| Grade III  | 49 (39%)       | 37 (29%)     |
| Grade IV   | 25 (20%)       | 42 (33%)     |
| OA at baseline                                     |                |              |
| Other knee affected                                | 109 (86%)      | 108 (84%)    |
| Other joints affected                              | 95 (75%)       | 87 (68%)     |

Values are mean (SD) or n (%).  
 \*Radiology grading is based on central grading, which may have differed from the site investigator's determination for patient eligibility.  
 AC, appropriate care; AC+H, appropriate care + hylan G-F 20; NSAID, non-steroidal anti-inflammatory drug; OA, osteoarthritis.

In the original study analysis,<sup>14</sup> the primary effectiveness measure was the mean change in the WOMAC pain subscore in the study knee from baseline to month 12. Secondary effectiveness measures included two definitions of a responder that incorporated a minimum clinically important response level of at least 20%. These measures were defined as the percentage of patients improved by month 12 (compared with baseline) using different combinations of the WOMAC subscales as follows: (1) at least a 20% improvement from baseline in the WOMAC pain score in the study knee (WOMAC 20P); (2) at least a 20% improvement from baseline in the WOMAC pain score in the study knee *and* at least a 20% improvement from baseline in either the function score or the stiffness score (WOMAC 20PFS).

**Alternative patient responder criteria**

Alternative patient responder criteria were examined in this analysis. Recent trials in rheumatoid arthritis have used higher threshold levels to define a patient responder, to "raise the bar" and define rheumatoid arthritis improvements by more substantial changes in core set measures.<sup>7</sup> While the 20% minimum clinically important response level used to define a patient responder in our original study was able to discriminate between the AC+H and AC treatment groups, we increased the minimum clinically important response levels to 50% and 70%.

These new criteria incorporate the pain, function, and stiffness subscores from the WOMAC, identical to the original secondary effectiveness measures. For the 50% minimum clinically important response level, the definitions were: (1) at least a 50% improvement from baseline in the WOMAC pain score in the study knee (WOMAC 50P); and (2) at least a 50% improvement from baseline in the WOMAC pain score in the study knee *and* at least a 50% improvement from baseline in either the function score or the stiffness score (WOMAC 50PFS). Similarly, for the 70% minimum clinically important response level, the definitions were: (1) at least a 70% improvement from baseline in the WOMAC pain score in the study knee (WOMAC 70P); and (2) at least a 70% improvement from baseline in the WOMAC pain score in the study knee *and* at least a 70% improvement from baseline in either the function score or the stiffness score (WOMAC 70PFS). These responder criteria can be collectively termed the WOMAC 20, WOMAC 50, and WOMAC 70 criteria.

**Differences between treatment groups**

Discriminant validity, which has been defined as the ability of a measure to distinguish clinically important differences between treatment groups,<sup>17</sup> was evaluated using these responder criteria. We hypothesised that when increasing the threshold for defining patient improvement, the number of patients classified as responders in both treatment groups would decrease. However, it is unclear how this would affect

**Table 2** WOMAC scores at baseline

| Baseline measure          | AC+H (n = 127) | AC (n = 127)* |
|---------------------------|----------------|---------------|
| WOMAC pain (0 to 20)†     | 11.4 (2.7)     | 11.9 (2.9)    |
| WOMAC stiffness (0 to 8)† | 5.1 (1.5)      | 5.1 (1.4)     |
| WOMAC function (0 to 68)† | 39.5 (9.3)     | 40.2 (9.3)    |

Values are mean (SD).  
 \*One patient in the AC group did not have a baseline WOMAC questionnaire completed and was thus not included in the analysis.  
 †The higher the score, the worse the problem.  
 AC, appropriate care; AC+H, appropriate care + hylan G-F 20; WOMAC, Western Ontario and McMaster Universities osteoarthritis index.

the overall treatment group differences for each patient improved definition.

In the original study, a 20% difference between treatment groups for the primary and secondary effectiveness measures was established a priori by the steering committee as the minimum clinically important difference based in part on previous research.<sup>18</sup> In addition, a 20% improvement was the minimum clinically important improvement from baseline to month 12 for each patient who was classified as a responder.

### Statistical methods

Data from the locked study database were analysed using SAS version 8.2. Multivariable logistic analyses were undertaken for each of the responder criteria that incorporated different minimum clinically important response levels. Patient were classified responders if they improved according to the criteria outlined in the definition from baseline to month 12. The hypothesis tested was whether AC+H was superior to AC when the responder criteria were applied.

All analyses were adjusted for design variables—that is, baseline value of the variable being analysed, site, blocking by site, body mass index, and baseline WOMAC total score. The type 1 experiment-wise error rate was controlled for by distributing  $\alpha$  over all six response levels (that is, WOMAC 20P and 20PFS; WOMAC 50P and 50PFS; WOMAC 70P and 70PFS) using the Bonferroni adjustment of  $\alpha/6$  ( $\alpha$  for each comparison =  $0.05/6 = 0.0083$ ). The original secondary effectiveness measures are provided for comparison with the patient improved definitions which incorporate higher minimum clinically important response levels.

All patients were included in the intent to treat group for all analyses as described earlier.<sup>14</sup> The hot deck method was used to impute data for all effectiveness measures as described earlier.<sup>14</sup>

## RESULTS

### Patients

In the trial, 128 patients were randomised to receive appropriate care and 127 patients to receive appropriate care with hylan G-F 20. In all, 24 patients dropped out of the study (21 in the AC group, three in the AC+H group). One patient in the AC group did not have a baseline WOMAC questionnaire completed and thus was not included in the analysis. Descriptive statistics comparing demographic variables, baseline disease characteristics, and baseline outcome measures (that is, WOMAC pain, function, and stiffness subscores) are given in tables 1 and 2. Overall, treatment groups were similar for demographics, disease characteristics,

and osteoarthritis treatments used at baseline. However, 20% of patients in the AC+H group and 33% in the AC group had grade IV osteoarthritis, as subsequently determined by central radiological grading. WOMAC scores for pain, stiffness, and function were similar between groups.

### Knee osteoarthritis treatment

All patients except one in the AC+H group had at least one course of hylan G-F 20 in their study knee, and 53 (42%) had at least one course in their contralateral knee. Forty five patients (38%) in the AC+H group received a second course of hylan G-F 20 in their study knee, and three received a third course in their study knee. Twenty patients (16%) in the AC+H group received a second course in their contralateral knee. More patients in the AC group than in the AC+H group reported corticosteroid injections in the study knee (70% v 14%) or in the contralateral knee (27% v 6%) (both  $p < 0.0001$ ). There were also more patients in the AC group taking NSAIDs for any knee (79% v 65%) ( $p = 0.0062$ ), and other drugs (20% v 10%) (for example, opioid analgesics, anti-inflammatory agents) for any knee ( $p = 0.0216$ ). There were no significant differences between the groups in the use of concomitant drug treatment for overall osteoarthritis. Further details of the knee osteoarthritis treatment can be found in the original study results.<sup>14</sup>

### Effectiveness

The results for the original secondary effectiveness measures and new responder criteria are given in table 3. They showed that for both the original secondary effectiveness measure and the alternative patient responder criteria, the percentage of responders was greater in the AC+H group than in the AC group. The treatment group differences were significant at the 0.0083 level ( $\alpha/6 = 0.05/6$ ) for the 20% and 50% minimum clinically important response levels (adjusted using Bonferroni correction) and exceeded the required 20% difference established a priori as the minimum clinically important difference between treatment groups (25% to 29%). When the minimum clinically important response level increased to 70%, the treatment group differences were approximately one half the size (that is, 11% to 12%) of the differences found with the 20% and 50% levels, and did not reach statistical significance after Bonferroni correction. Within each minimum clinically important response level, the treatment group differences were similar regardless of whether the WOMAC pain scores, or all of the WOMAC pain, function, and stiffness scores, were incorporated into the

**Table 3** Number (%) of patient responders using WOMAC 20, 50, and 70 minimum clinically important response levels

| Patient responder definitions          | Treatment group |               | Difference    |          |
|--|-----------------|---------------|---------------|----------|
|  | AC+H (n = 127)  | AC* (n = 127) | [(AC+H) – AC] | p Value  |
| <i>Original effectiveness measures</i> |                 |               |               |          |
| WOMAC 20P†                             | 87 (69%)        | 51 (40%)      | 29%           | 0.0001‡  |
| WOMAC 20PFS†                           | 79 (62%)        | 45 (35%)      | 27%           | 0.0001‡  |
| <i>Alternative responder criteria</i>  |                 |               |               |          |
| WOMAC 50P                              | 53 (42%)        | 20 (16%)      | 26%           | <0.0001‡ |
| WOMAC 50PFS                            | 43 (34%)        | 12 (9%)       | 25%           | <0.0001‡ |
| WOMAC 70P                              | 26 (20%)        | 10 (8%)       | 12%           | 0.0118   |
| WOMAC 70PFS                            | 20 (16%)        | 6 (5%)        | 11%           | 0.0100   |

\*One patient in the AC group did not have a baseline WOMAC questionnaire completed and thus was not included in the analysis.

†Secondary effectiveness measures from main study.

‡Analysis controlled for the type I experiment-wise error rate by distributing  $\alpha$  over all response levels using Bonferroni adjustment  $\alpha/6 = 0.05/6 = 0.0083$ .

AC, appropriate care; AC+H = appropriate care + hylan G-F 20.

definition (for example, 29% for WOMAC 20P, 27% for WOMAC 20PFS).

The percentage of patients classified as responders decreased for both treatment groups as response levels increased from 20% to 70%, and with the more stringent definition incorporating pain, function, and stiffness within each response level. Considering the AC+H group, when moving from the lower to the higher response level for pain only (that is, WOMAC 20P to WOMAC 70P), the percentage of responders decreased from 69% to 20%. Similarly, when increasing the response levels with the more stringent criteria incorporating pain and either function or stiffness (that is, WOMAC 20PFS to WOMAC 70PFS), a similar decrease was observed in the AC+H group (62% to 16%). For the AC group, large decreases were also found when response levels increased for the criteria incorporating only pain (40% to 8%), and the more stringent criteria incorporating pain and either function or stiffness (35% to 5%).

When comparing the AC+H group and the AC group for all responder criteria, the results show that the percentage of responders in the AC+H group relative to the AC group was generally greater for criteria that incorporate the higher minimum clinically important response levels. For example, for the WOMAC 70PFS criterion, the percentage of responders in the AC+H group was approximately three times the percentage of responders in the AC group (that is, 16% *v* 5%). This is in comparison to the WOMAC 20PFS criterion where the percentage of responders in the AC+H group was less than twice the percentage of responders in the AC group (that is, 62% *v* 35%). This pattern was also observed with the criteria incorporating pain (WOMAC 20P to WOMAC 70P).

## DISCUSSION

Traditional methods of carrying out between group comparisons of clinical trials data are often based on the analysis of continuous variables. These provide an appreciation of the magnitude and variation of group effects but do not usually translate into an understanding of the degree of improvement experienced by individual patients. In contrast, responder criteria, while being reductionist from a group standpoint, are capable of categorising individual patients according to whether they achieve levels of improvement at or above prespecified response thresholds. Response thresholds have generally been established a priori either to reflect a clinically important difference at an individual level, or on the basis of differentiating most efficiently between an active treatment and a placebo control.<sup>4-6</sup> In the case of the effectiveness measures used in the original study,<sup>14</sup> these were proposed during protocol development at a time when there was no precedent to follow, but 20% was considered by the development group to represent a minimum clinically important difference, and one that was of the same order of magnitude as the previously published ACR 20 criteria<sup>4</sup> for rheumatoid arthritis. The OARSI responder criteria<sup>5</sup> were developed during the execution of the protocol, and the OMERACT-OARSI responder criteria<sup>6</sup> were developed following completion of the study, but neither were available at study initiation. It is of interest therefore that WOMAC 20 and WOMAC 50 responder criteria, based on pain only or on the pain, stiffness, and function subscales, yield statistically detectable between group differences of the order of 25% to 29%, with percentage response at WOMAC 20 being slightly higher numerically than at WOMAC 50. Indeed this approach to the analysis provides additional confirmation of the clinical and statistical superiority of adding hylan G-F 20 to appropriate care regimens in the treatment of knee osteoarthritis. While differences at the highest threshold level (WOMAC 70) were not statistically detectable after Bonferroni correction and may be more difficult to attain,

an appropriately powered study could be capable of detecting differences in patient attainment rates at even this very high level of percentage improvement. This approach to dissecting the differential therapeutic response can be considered complementary to other responder criteria and should not be considered as replacing more traditional methods. Whether these observations can be generalised to patients with either more or less severe symptoms requires further study.

A potential limitation of response criteria based on percentage change is that the accompanying absolute change can differ markedly. Thus a 20% improvement for a patient with a baseline score of 20 normalised units (NU) (0–100 NU scale) is 4 units, whereas a 20% improvement for a patient with a baseline score of 75 NU is 15 units. Furthermore, in a comparison of outcome measures in rheumatoid arthritis clinical trials, Anderson *et al*<sup>19</sup> noted that measures based on continuous data provided better responsiveness than the ACR 20 or disease activity score. Nevertheless, response criteria based on percentage change offer simplicity, some comparability with OMERACT-ACR criteria for rheumatoid arthritis, and an opportunity to review the number of patients who attain or exceed a prespecified threshold. While the use of responder criteria may have a negative impact on statistical power for clinical trials applications, it does provide a novel approach to outcome measurement which may facilitate the use of quantitative measurement procedures in clinical practice applications.

The results of this analysis provide evidence for the capacity of WOMAC 20, 50, and 70 responder criteria to detect clinically important and statistically significant differences between two active treatment groups in a pragmatic randomised trial. In particular we have observed—as judged by each of the four criteria sets with Bonferroni correction and by all six criteria sets without correction—that significantly more patients in the AC+H group achieved responder status than in the AC group. This approach, based on percentage improvement in pain alone or in pain and either stiffness or function, allows reviewers and consumers to discern how many patients experienced a clinically important reduction in symptom severity. Given that the analytical strategy is individualised, this approach may have important implications for monitoring patients in routine clinical care and facilitating evidence based therapeutic decision making and shared goal setting in various health care environments.

## ACKNOWLEDGEMENTS

We thank Genzyme Corporation for funding the study. The study agreement with Genzyme Corporation gave the investigators independence to publish regardless of the results. We would also like to thank the clinical investigators who enrolled patients and are listed in the main manuscript.

## Authors' affiliations

**N Bellamy**, CONROD, Faculty of Health Sciences, The University of Queensland, Brisbane, Queensland, Australia

**M J Bell**, Rheumatology and Clinical Epidemiology, Sunnybrook & Women's Health Sciences Centre, University of Toronto, Toronto, Ontario, Canada

**C H Goldsmith**, **G W Torrance**, Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada

**D Pericak**, **V Walker**, Innovus Research Inc, Burlington, Ontario, Canada

**J P Raynauld**, Department of Medicine, Montreal Institute of Rheumatology, Montreal, Quebec, Canada

**P Tugwell**, Centre for Global Health, Institute of Population Health, University of Ottawa, Ottawa, Ontario, Canada

**R Polisson**, Genzyme Corporation, Cambridge, Massachusetts, USA

## REFERENCES

- 1 **Felson DT**, Anderson JJ, Boers M, Bombardier C, Chernoff M, Fried B, *et al*. The American College of Rheumatology preliminary core set of disease activity measures for rheumatoid arthritis clinical trials. The Committee on Outcome Measures in Rheumatoid Arthritis Clinical Trials. *Arthritis Rheum* 1993;**36**:729–40.
- 2 **Osteoarthritis Research Society (OARS) Task Force**. Design and conduct of clinical trials of patients with osteoarthritis: recommendations from a task force of the Osteoarthritis Research Society. *Osteoarthritis Cart* 1996;**4**:217–43.
- 3 **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 in OMERACT III. *J Rheumatol* 1997;**24**:799–802.
- 4 **Felson DT**, Anderson JJ, Boers M, Bombardier C, Furst D, Goldsmith CH, *et al*. American College of Rheumatology. Preliminary definition of improvement in rheumatoid arthritis. *Arthritis Rheum* 1995;**38**:727–35.
- 5 **Dougados M**, LeClaire P, van der Heijde D, Bloch DA, Bellamy N, Altman RD, *et al*. Response criteria for clinical trials on osteoarthritis of the knee and hip: a report of the Osteoarthritis Research Society International Standing Committee for Clinical Trials Response Criteria Initiative. *Osteoarthritis Cart* 2000;**8**:395–403.
- 6 **Pham T**, van der Heijde D, Lassere M, Altman RD, Anderson JJ, Bellamy N, *et al*. Outcome variables for osteoarthritis clinical trials: the OMERACT-OARSI set of responder criteria. *J Rheumatol* 2003;**30**:1648–54.
- 7 **Felson DT**, Anderson JJ, Lange ML, Wells G, 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.
- 8 **Kalden JR**, Schattenkirchner M, Sorensen H, Emery P, Deighton C, Rozman B, *et al*. The efficacy and safety of leflunomide in patients with active rheumatoid arthritis: a five-year follow up study. *Arthritis Rheum* 2003;**48**:1513–20.
- 9 **Hochberg MC**, Tracy JK, Hawkins-Holt M, Flores RH. Comparison of the efficacy of the tumour necrosis factor alpha blocking agents adalimumab in rheumatoid arthritis. *Ann Rheum Dis* 2003;**62**(suppl 2):13–16.
- 10 **Kremer JM**, Weinblatt ME, Bankhurst AD, Bulpitt KJ, Fleischmann RM, Jackson CG, *et al*. Etanercept added to background methotrexate therapy in patients with rheumatoid arthritis: continued observations. *Arthritis Rheum* 2003;**48**:1493–9.
- 11 **Cohen S**, Hurd E, Cush J, Schiff M, Weinblatt ME, Moreland LW. Treatment of rheumatoid arthritis with anakinra, a recombinant human interleukin-1 receptor antagonist, in combination with methotrexate: results of a twenty-four week, multicenter, randomized, double-blind, placebo-controlled trial. *Arthritis Rheum* 2002;**46**:614–24.
- 12 **Weisman MH**, Moreland LW, Furst DE, Weinblatt ME, Keystone EC, Paulus HE, *et al*. Efficacy, pharmacokinetic, and safety assessment of adalimumab, a fully human anti-tumor necrosis factor-alpha monoclonal antibody, in adults with rheumatoid arthritis receiving concomitant methotrexate: a pilot study. *Clin Ther* 2003;**25**:1700–21.
- 13 **Mease PJ**, Goffe BS, Metz J, VanderStoep A, Finck B, Burge DJ. Etanercept in the treatment of psoriatic arthritis and psoriasis: a randomized trial. *Lancet* 2000;**356**:385–90.
- 14 **Raynauld JP**, Torrance GW, Band PA, Goldsmith CH, Tugwell P, Walker V, *et al*. A prospective, randomized, pragmatic health outcomes trial evaluating the incorporation of hylan G-F 20 into the treatment paradigm for patients with knee OA: clinical results. *Osteoarthritis Cart* 2002;**10**:506–17.
- 15 **Torrance G**, Raynaud JP, Walker V, Goldsmith CH, Bellamy N, Band P, *et al*. A prospective randomized health outcomes trial evaluating the incorporation of hylan GF-20 into the treatment paradigm for patients with knee osteoarthritis: economic results. *Osteoarthritis Cart* 2002;**10**:518–27.
- 16 **Hochberg MC**, Altman RD, Brandt KD, Clark BM, Dieppe PA, Griffin MR, *et al*. Guidelines for the medical management of osteoarthritis. Part II. Osteoarthritis of the knee. *Arthritis Rheum* 1995;**38**:1541–6.
- 17 **Bombardier C**, Tugwell P. A methodological framework to develop and select indices for clinical tasks: statistical and judgmental approaches. *J Rheumatol* 1982;**9**:753–7.
- 18 **Goldsmith CH**, Boers M, Bombardier C, Tugwell P. Criteria for clinically important changes in outcomes: development, scoring, and evaluation of rheumatoid arthritis patient and trial profiles. OMERACT Committee. *J Rheumatol* 1993;**20**:561–5.
- 19 **Anderson JJ**, Bolognese JA, Felson DT. Comparison of rheumatoid arthritis clinical trial outcome measures. *Arthritis Rheum* 2003;**48**:3031–8.