Outcome Variables for Osteoarthritis Clinical Trials: The OMERACT-OARSI Set of Responder Criteria

THAO PHAM, DÉSIRÉE VAN DER HEIJDE, MARISSA LASSERE, ROY D. ALTMAN, JENNIFER J. ANDERSON, NICHOLAS BELLAMY, MARC HOCHBERG, LEE SIMON, VIBEKE STRAND, THASIA WOODWORTH, and MAXIME DOUGADOS

ABSTRACT. Improvement in analysis and reporting results of osteoarthritis (OA) clinical trials has been recently obtained because of harmonization and standardization of the selection of outcome variables (OMERACT3 and OARSI). Moreover, OARSI has recently proposed the OARSI responder criteria. This composite index permits presentation of results of symptom modifying clinical trials in OA based on individual patient responses (responder yes/no). The 2 organizations (OMERACT and OARSI) established a task force aimed at evaluating: (1) the variability of observed placebo and active treatment effects using the OARSI responder criteria; and (2) the possibility of proposing a simplified set of criteria. The conclusions of the task force were presented and discussed during the OMERACT 6 conference, where a simplified set of responder criteria (OMERACT-OARSI set of criteria) was proposed. (J Rheumatol 2003;30:1648–54)

Key Indexing Terms:
OSTEOARTHRITIS OUTCOMES CLINICALRESPONDER CRITERIA

The approach to the conduct of clinical trials in patients with osteoarthritis (OA) has evolved over the past decades[1], and significant progress has been made in outcome measurement procedures for OA clinical trials. However, trials in OA treatments report the average response in multiple outcome measures for treated patients, and further international harmonization of measurement procedures used to establish beneficial effects in phase III clinical trials is required[2].

Which patients are to be considered in a phase III OA clinical trial?

The diagnosis of OA is usually based primarily on radiographic rather than clinical features[3]. However, efficacy in OA clinical trials is determined by improvement in symptoms (pain, function, stiffness), in WOMAC questionnaire score (Western Ontario McMaster University OA Index), and global evaluation. The American Rheumatism Association criteria for diagnosis of OA, a classification rather than a set of diagnostic criteria, are accurate for reporting series of cases, because they are consistent and improve communication. Nevertheless, patients entering phase III symptom modifying trials should fulfill these criteria, and, as recommended by the task force of the Osteoarthritis Research Society (OARSI), should also have pain of at least mild intensity[4]. The task force also indicated that patients entering trials of disease modifying agents should either be free of OA if the agent is being evaluated for a preventative effect, or have mild to moderate OA, if the agent is being evaluated for effects on the rate of progression. Presence of symptoms is not mandatory for entry into a trial of a potential disease modifying agent (so-called DMOAD or joint structure preserving clinical trials). This emphasizes the importance of the selection and standardization of outcome variables.

Which outcome variables are to be considered in the evaluation of symptoms in OA?

In the evaluation of symptoms of OA, several domains can be considered. The OARSI task force proposed that pain should be the primary outcome variable in trials of OA agents based on symptoms[4]. This approach was based on the recommendations of the OMERACT 3 conference, which was focused on identifying a core set of measures to be considered for future phase III clinical trials[2]. A consensus was reached on 3 symptomatic domains that should system-

From René Descartes University, AP-HP, Rheumatology B Department, Cochin Hospital, Paris, France.

T. Pham, MD, René Descartes University, Cochin Hospital; Aix-Marseille II University, Conception Hospital, Marseille, France; D. van der Heijde, MD, PhD, Professor, Department of Rheumatology, University Hospital and Biomedical Research Institute, Maastricht, The Netherlands; M.N.D. Lassere, MB, BS, PhD, FRACP, FAFPHM, Staff Specialist in Rheumatology, Department of Rheumatology, St. George Hospital, University of NSW, Sydney, Australia; R.D. Altman, MD, University of Miami, Department of Veterans Affairs Medical Center, Miami, FL; J.J. Anderson, PhD, Clinical Epidemiology Research and Training Unit, Boston Medical Center, Boston, MA, USA; N. Bellamy, MD, MSc, FACP, FRCP, FRCPC, FAFRM, FRACP, Department of Medicine, University of Queensland, Royal Brisbane Hospital, Queensland, Australia; M. Hochberg, MD, MPH, Division of Rheumatology and Clinical Immunology, Department of Medicine, University of Maryland School of Medicine, Baltimore, MD; L. Simon, MD, Department of Medicine, Beth Israel Deaconess Medical Center, Boston, MA; V. Strand, MD, Biopharmaceutical Consultant, Portola Valley, CA; T. Woodworth, Pfizer Global Research and Development, New London, CT, USA; M. Dougados, MD, René Descartes University, Cochin Hospital, Paris, France.

Address reprint requests to Dr. M. Dougados, Rheumatology B Department, Cochin Hospital, 27 rue du Faubourg Saint Jacques, 75014 Paris, France.

atically be included in all future phase III clinical OAtrials: pain, physical function, and patient global assessment. Consensus followed a process based on 4 elements, i.e., provision of information from the literature, lectures followed by a discussion period, breakout groups, and polling procedures. This consensus process did not exclude the possible addition of other measures to the core set, such as physician global assessment, generic quality of life/utility, inflammation, stiffness, or time to surgery.

Within each domain, several instruments may be considered. For use in future phase III clinical trials, these instruments should have adequate reliability, validity, and responsiveness. They may be simple, like the visual analog scale (VAS)[5] or the Likert scale, or more complex (e.g., a questionnaire). For instance, the 5-item pain scale included in the WOMAC[6] and the Lequesne Functional Severity Index[7] are widely used in clinical trials; these instruments have been shown to be valid, reliable, and sensitive to change in patients with OAof the hip and/or the knee.

Why a composite index that includes outcome variables selected for future phase III OA trials?

Several arguments favor presenting results of the effect of therapy on an individual basis, e.g., to evaluate whether an individual patient responds to a treatment or not. This approach has been used successfully in rheumatoid arthritis[8]. In OA, mean changes of multiple dimensions (pain, function, range of motion, etc.) are currently used in clinical trials. A composite index that includes information from various domains presented as a dichotomous (yes/no) variable has several advantages. It has greater face validity, uses simple and consistent medical language, and simplifies statistical analysis and avoids multiple tests.

The goal of the OARSI standing committee for clinical trials response criteria was to promulgate a single definition for use in future OAtrials[9] following a data driven approach. For this purpose, information was obtained by analyses of 14 OA symptom modifying clinical trials in order to test the optimal combinations of different symptomatic outcome variables. Three domains selected by the OMERACT 3 process were included in the composite index: pain, function, and patient global assessment. Whatever instrument was used, it had to be evaluated using a 0–100 scale, and changes observed during study had to be evaluated using the LOCF (last observation carried forward) technique. Based on this data driven and expert opinion approach, 2 propositions were made: Proposition A defined a responder as having achieved a high degree of improvement in pain or a moderate degree of improvement in 2 of the 3 response domains (pain, function, global assessment). Proposition B defined a responder as having achieved a high degree of improvement in either pain or function, or a moderate degree of improvement in 2 of the 3 response domains (Figure 1). These propositions, presented as a combination of a tree-format and a matching-format, incorporated both relative and absolute changes, and different cut-offs with regard to the evaluated drug, the OA localization, and the domain (pain, function, patient global assessment).

The formal OARSI criteria outcome variable domains were those agreed upon during the OMERACT 3 conference. They were highly correlated to patient self–report domains, and one can regret the absence of physician global report and laboratory-based domains. However, these domains could be included in future OA trials as secondary outcome variables. The formal OARSI criteria format optimized discrimination and increased power. However, there was a loss of simplicity because of the multiple cut-offs with regard to the evaluated drug, the OA localization, and the domain. Additional cut-offs were examined in an effort to determine a uniform cut-off for all subsets, but those tested showed a clear loss of sensitivity and specificity with the database used[9].

The use of such composite indices is still controversial, however. Apotential limitation of the OARSI criteria is that the task force did not use specific methods to avoid multiplicity, such as O'Brien's nonparametric rank-sum procedure, in the development of the criteria[10]. Tilley[11] showed that O'Brien's global statistic, comprising components of the American College of Rheumatology 20% response criteria (ACR20), was more precise than the binary ACR20 responses or any of its individual components. Using data from trials of nonsteroidal antiinflammatory drugs (NSAID) in patients with hip or knee OA, Bolognese, *et al* demonstrated that individual study endpoints were highly correlated and concluded that composite endpoints did not increase the precision of efficacy comparisons beyond that of individual endpoints[12]. Therefore, the OMERACT-OARSI task force reviewed the formal OARSI responder criteria to determine whether a simplified set of criteria could be developed.

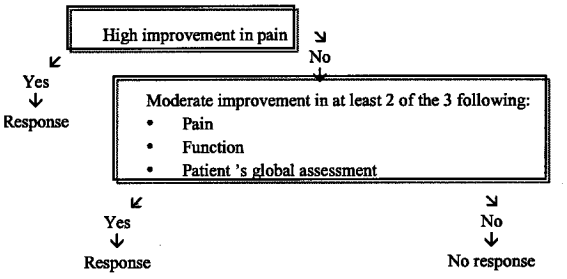
OMERACT-OARSI Initiative

Rationale. OMERACT and OARSI proposed a task force aimed at evaluating the following points:

1. Variability of the observed placebo and active treatment effects when using the formal OARSI set of criteria, in different databases
2. Whether the use of a simplified set of criteria might be proposed

Comparison of "elaboration" versus "revisit" databases. Using the formal OARSI responder criteria, performance between 2 different databases were compared. Two databases developed from clinical double-blind, randomized, placebo controlled trials were used for this data driven approach:

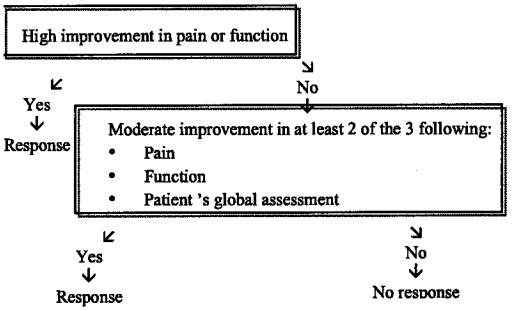
• The elaboration database, which served for the development of the formal OARSI criteria (1886 patients included in 14 OAknee or hip clinical trials with a duration of at least 6 weeks)



Proposition A

Subgroup	High improvement in pain				Moderate improvement in Pain		Moderate improvement in Function	
	Relative change*	Absolute change**	Relative change	Absolute change	Relative change	Absolute change	Relative change	Absolute change
Knee oral NSAID	45	20	15	10	30	15	35	10
Knee oral specific drug	55	30	35	10	15	20	15	15
The 3 above groups together	55	30	35	15	15	20	15	15
Knee IA specific drug	40	30	35	15	35	10	30	10

Optimal cut-offs to be applied for the OARSI Responder Criteria



Proposition B

Subgroup	High improvement in Pain		High improvement in Function		Moderate improvement in Pain		Moderate improvement in Function		Moderate improvement in Global assessment	
	Relative change*	Absolute change**	Relative change	Absolute change	Relative change	Absolute change	Relative change	Absolute change	Relative change	Absolute change
Hip NSAID	50	30	50	20	25	15	20	10	20	10
Knee oral NSAID	50	20	60	20	30	15	20	20	25	10
Knee oral specific drug	55	30	50	20	30	20	20	20	20	15
The 3 above groups together	55	30	50	20	30	15	20	20	20	15
Knee IA specific drug	50	30	60	20	20	20	30	10	30	10

Optimal cut-offs to be applied for the OARSI Responder Criteria

Figure 1. OARSI formal set of criteria: Scenarios A and B. *Relative change: percentage of change during the study (final– baseline / baseline × 100). IA: intraarticular. **Absolute change: absolute change during the study (final–baseline on a 0–100 scale).

• The revisit database (15 trials involving 8164 patients included in OA knee or hip clinical trials with a duration of at least 4 weeks). Flow diagrams of the 2 databases are summarized in Figure 2.

The data included in the 2 databases were from individual positive OAclinical trials that had demonstrated clinical benefit and included a placebo treatment group. The analyses were conducted according to a protocol provided to each of the contributing pharmaceutical companies. The definition of a "positive" trial was based on a p value < 0.05 for the *a priori* chosen primary criterion of the trial. Only intention-to-treat analysis trials using the LOCF technique

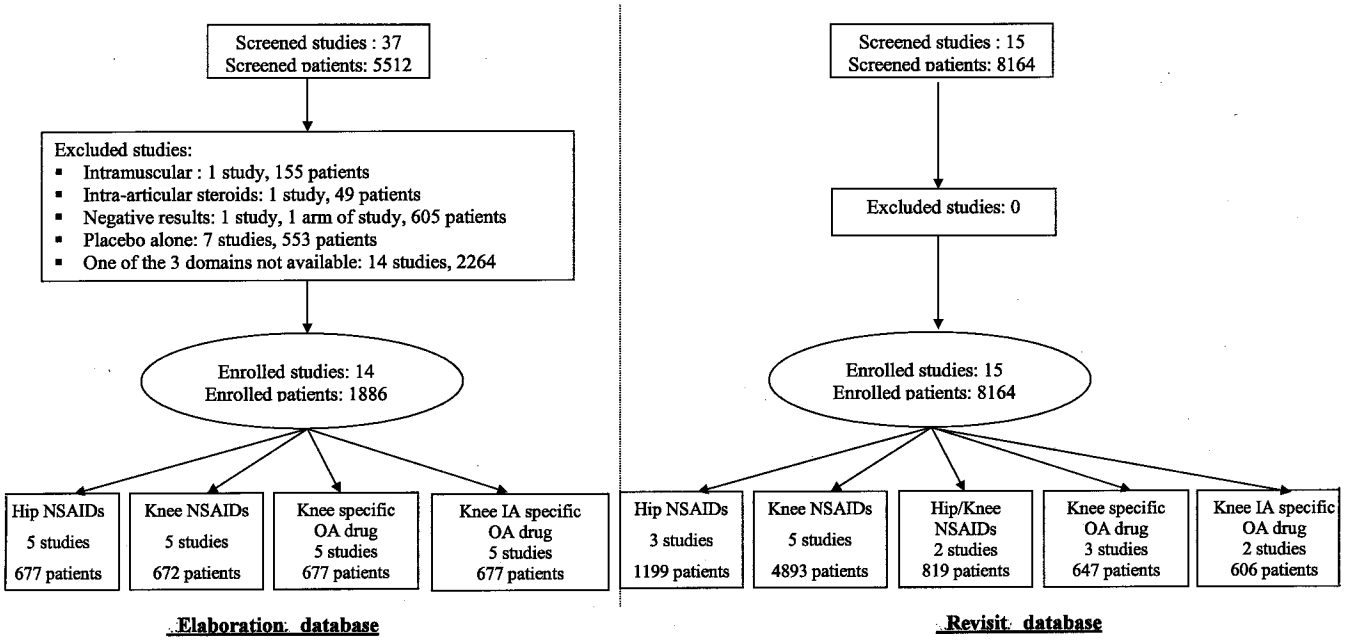


Figure 2. Flow diagrams of the 2 databases. IA: intraarticular.

were used. In both databases, the drugs were not identified by name but by class of agent; the majority of the information concerned NSAID in hip and knee OA.

For the revisit database the drug companies provided, for each scenario, the percentage of patients receiving active drug who responded according to the different scenarios, and the percentage of patients receiving inactive drug (placebo) who were non-responders; individual endpoints were not provided.

The first step of the current study consisted of the evaluation of performance between these 2 databases using 2 formal propositions labeled in this article as "scenario A" and "scenario B." That is, we compared results in the elaboration database with those from analyses conducted with the revisit database, and further characterized the following performance: the placebo effect, i.e., the percentage of responders in the placebo group; the active effect, i.e., the percentage of responders in the active treatment group; the treatment effect, i.e., the percentage of patients improved in the active treatment group minus the percentage of patients improved in the placebo group and the sample size needed to obtain the placebo effect and the active treatment effect observed ($\alpha = 0.05$ and $\beta = 0.20$, 2 tailed test).

Results are summarized in Table 1. For both scenarios (A and B), the difference in placebo and active treatment effects was quite large between the 2 databases (4% to 21% for placebo effect, 7% to 34% for active treatment effect, and 4% to 22% for treatment effect). Based on the results, and considering the placebo and active treatment effects in the elaboration database, the sample size required in future NSAID trials in knee OA was calculated to be 67 patients per arm with scenario A and 66 with scenario B, using the responder criteria.

Development of the OMERACT-OARSI Responder Criteria

The second step of the ongoing study was the development of OMERACT-OARSI responder criteria for OA clinical phase III trials, using both a data driven approach and an expert opinion approach.

In the data driven approach, the performance observed using 6 scenarios of different sets of responder criteria were compared. The first 2 scenarios were the 2 propositions (A and B) of the formal OARSI set of criteria[5] (Figure 1). The 4 other scenarios (scenarios C to F) were proposed by the OMERACT scientific committee. Their main feature was use of a uniform cut-off, irrespective of OA localization, study drug, or route of administration; unlike the formal OARSI criteria (Figure 3). Scenarios A, C, and E, defined response primarily as pain relief (high improvement), while in scenarios B, D, and F it was pain or function. Scenarios C and D, like the formal OARSI criteria, defined response as relative change (percentage of change during the study) and absolute change (absolute change during the study), while scenarios E and F defined response in terms of only relative change.

These scenarios were then applied for analyses of the revisit database, and performance was considered relative to placebo effect, active treatment effect, treatment effect, and sample size required in future trials to obtain the observed placebo. Results are summarized in Table 2. In knee OA NSAID trials, the highest active treatment effect was observed when using scenario F (66.4%) and the lowest

Table 1. Performances observed with the formal sets of criteria, proposition A and B *(e.g., scenarios A and B) in the elaboration and in the revisit databases: placebo effect, active treatment effect, treatment effect, and variability between the 2 databases.

	Formal OARSI Set of Criteria					
	Proposition A (Pain)*			Proposition B (Pain or Function)*		
Trials	Elaboration, %	Revisit, %	(Revisit–Elaboration)	Elaboration, %	Revisit, %	(Revisit–Elaboration)
Knee OA, systemic specific OAdrug						
Placebo effect	51	31	–20	50	29	–21
Active treatment effect	62	38	–24	61	36	–25
Treatment effect	11	7	–4	11	7	–4
Knee OAIASpecific OAdrug						
Placebo effect	47	35	–12	47	35	–12
Active treatment effect	92	58	–34	91	57	–34
Treatment effect	45	23	–22	44	22	–22
Hip OANSAIDs						
Placebo effect	33	29	–4	39	32	–7
Active treatment effect	62	54	–8	69	58	–11
Treatment effect	29	25	–4	30	26	–4
Knee OANSAIDs						
Placebo effect	27	39	+12	26	39	+13
Active treatment effect	52	59	+7	51	58	+7
Treatment effect	25	20	–5	25	19	–6

* See text for detailed explanations. Elaboration: database used to develop the formal sets of responder criteria[9] ; Revisit: database used to revisit the formal sets of responder criteria and to evaluate the simplified sets of responder criteria.

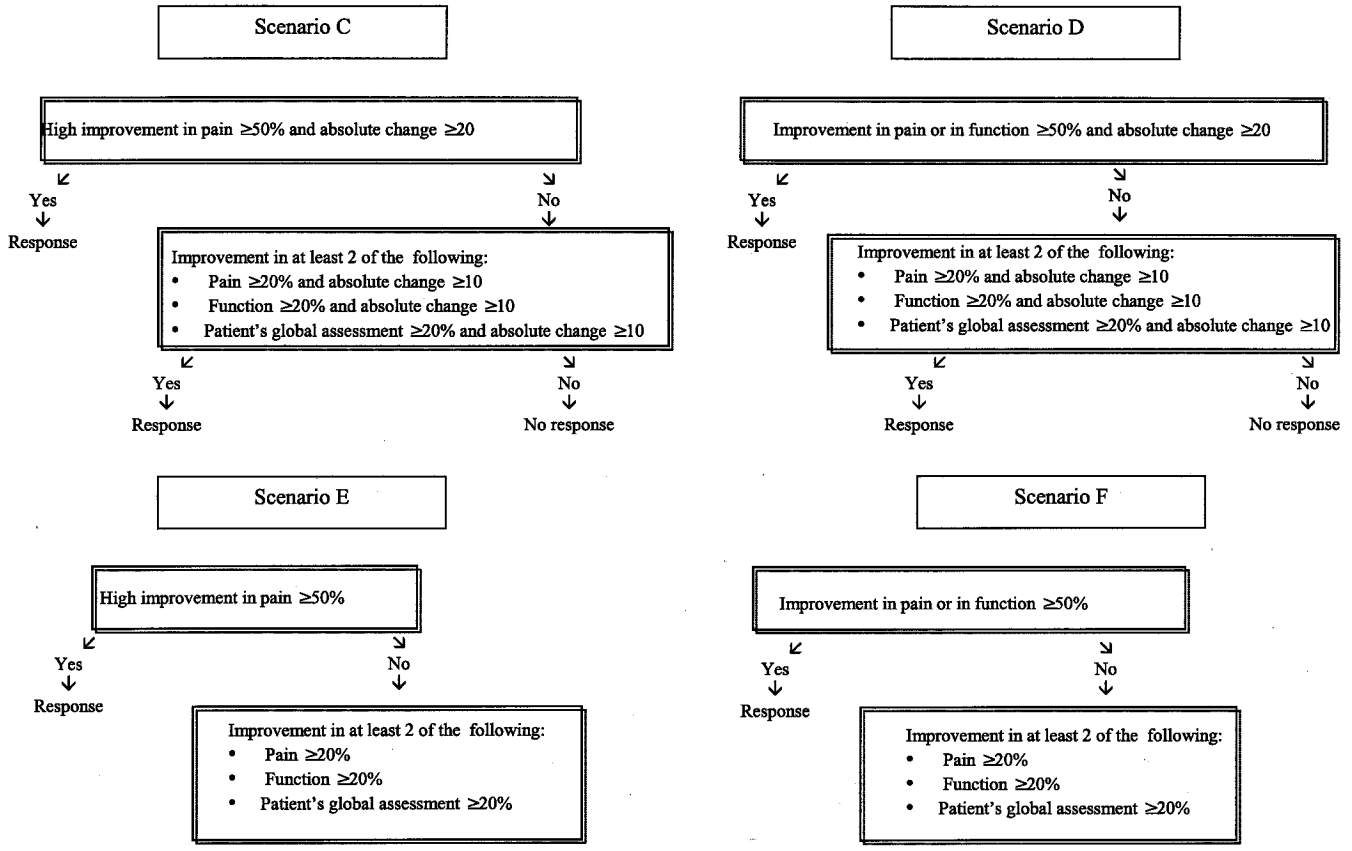


Figure 3. Evaluated simplified sets of criteria: Scenarios C to F.

placebo effect when using scenario B (39.1%). The treatment effect was similar whatever the scenario (19.8%, 19.3%, 19.8%, 19.5%, 19.9%, and 19.8% for scenarios A, B, C, D, E, and F, respectively). The sample size required in future NSAID trials in knee OA was also similar whatever the scenario (99, 105, 98, 101, 97, and 98 patients per arm for scenarios A, B, C, D, E, and F, respectively).

The results observed whatever the drug and whatever the localization were the following: the highest active treatment effect with scenario F (65.5%), and the lowest placebo effect with scenario B (35.6%). The treatment effect was similar whatever the scenario (21.5%, 21.9%, 21.7%, 21.3%, 20.4%, and 20.6% for scenarios A, B, C, D, E, and F, respectively). The sample size required in future trials in OAwas also similar whatever the scenario (84, 81, 82, 85, 93, and 91 patients per arm for scenarios A, B, C, D, E, and F, respectively).

Taking an "expert opinion approach," the results of the 2 steps of the current study were presented to OMERACT 6 conference participants (Brisbane 2002, OMERACT 6, OA workshop). Participants examined performance using the formal OARSI responder criteria between the 2 different databases (elaboration vs revisit), and then evaluated performance using the 6 different scenarios in the revisit database.

After presentation of the results, OMERACT 6 participants were asked to reply to several questions and an electronic polling system collected the responses. The following questions were asked:

• How important were the following characteristics (lowest placebo effect, highest active treatment effect, highest treatment effect, smallest sample size) for selecting the optimal set of responder criteria?

• Should a set of responder criteria in OAbe different with regard to OA localization (knee vs hip), drug (analgesics, NSAID, specific OA drugs), or route of administration (per os vs intraarticular)?

• With regard to results of the data driven approach and with regard to their previous responses, which scenario did they consider the most appropriate to be proposed as the OMERACT–OARSI set of responder criteria?

During the OMERACT 6 plenary session, results of the study and of the OA workshop vote were presented and the conference participants were asked to validate the choice of the OA workshop scenario. Results of the vote concerning the importance the OA workshop participants (n = 45) accorded (on a scale of 1 to 9) to each performance within an optimal set of responder criteria were: lowest placebo

Table 2. Performances observed with each scenario in the revisit database* : Percentage of patients improved in placebo and active treatment groups (i.e., placebo effect and active treatment effect), treatment effect, sample size per arm required in future trials, α = 0.05, ß = 0.20, two-tailed, expected placebo effect = that observed with this database, expected active treatment effect = that observed with this database.

	Knee OA, NSAID	Hip OA, NSAID	Any Joint, NSAID	Any Joint, Any Treatment
Scenario A				
Improved in placebo group, %	39.5	28.9	36.8	35.7
Improved in active group, %	59.3	53.6	58.3	57.2
Treatment effect, %	19.8	24.7	21.5	21.5
Sample size, n	99	62	84	84
Scenario B				
Improved in placebo group, %	39.1	31.5	37.4	35.6
Improved in active group, %	58.4	58.0	58.7	57.5
Treatment effect, %	19.3	26.5	21.3	21.9
Sample size, n	105	55	86	81
Scenario C				
Improved in placebo group, %	45.3	34.4	43.3	42.2
Improved in active group, %	65.1	60.3	64.7	63.9
Treatment effect, %	19.8	25.9	21.4	21.7
Sample size, n	98	58	84	82
Scenario D				
Improved in placebo group, %	45.9	34.8	43.9	42.9
Improved in active group, %	65.4	60.5	65.0	64.2
Treatment effect, %	19.5	25.7	21.1	21.3
Sample size, n	101	59	87	85
Scenario E				
Improved in placebo group, %	45.9	35.2	44.0	44.3
Improved in active group, %	65.8	60.5	65.3	64.7
Treatment effect, %	19.9	25.3	21.3	20.4
Sample size, n	97	61	85	93
Scenario F				
Improved in placebo group, %	46.6	35.5	44.5	44.9
Improved in active group, %	66.4	60.8	66.0	65.5
Treatment effect, %	19.8	25.3	21.5	20.6
Sample size, n	98	61	83	91

* Revisit database is that used to revisit the formal sets of responder criteria and to evaluate the simplified sets of responder criteria. ** Percentage of patients improved in the placebo group or in the active treatment group (i.e. placebo effect and active treatment effect).

effect (mean ± SD), 4.4 ± 4.07; highest active effect, 5.3 ± 5.37; treatment effect, 8.0 ± 8.5; and sample size required in future trials, 7.0 ± 4.15. OA workshop participants considered that a set of responder criteria in OA should not be different with regard to OAlocalization (62%), drug (67%), or route of administration (70%). And finally, OAworkshop participants considered that the most appropriate scenario to use as a set of OAresponder criteria was scenario D (45%); this choice was validated by the majority of participants [“yes, I agree” = 79% (58/73 votes)].

Even if the data driven approach did not allow selection of a particular scenario, the conclusion regarding this approach was that simplification of the set of criteria will not result in a loss of performance. In fact, a higher active treatment effect and a higher placebo effect were observed using the simplified scenarios. However, the treatment effect and the sample size required to obtain the placebo and active effects were similar whatever the scenario (simplified version of the formal or revised set).

These 2 outcomes of the data evaluation exercise were the most important for an optimal set of responder criteria according to the expert vote. Although similar results were observed with all evaluated scenarios, the experts chose scenario D (Figure 4), confirming the importance of a format that:

• requires an absolute change and a relative change, in order to assess a minimum entry criterion. Individuals with baseline values less than the minimum required for absolute change cannot be considered as responders.

• considers both pain and function as important domains, taking into account that in some studies, changes in functional disability are at least as important as changes in pain.

This set of criteria, endorsed by both the OARSI and the OMERACT committees, is at least as powerful as the previous OARSI formal set of criteria, and its simplification will probably facilitate its use in future OAtrials. However, the format of such a set of responder criteria implies a minimal level of symptoms at entry in order for the patient

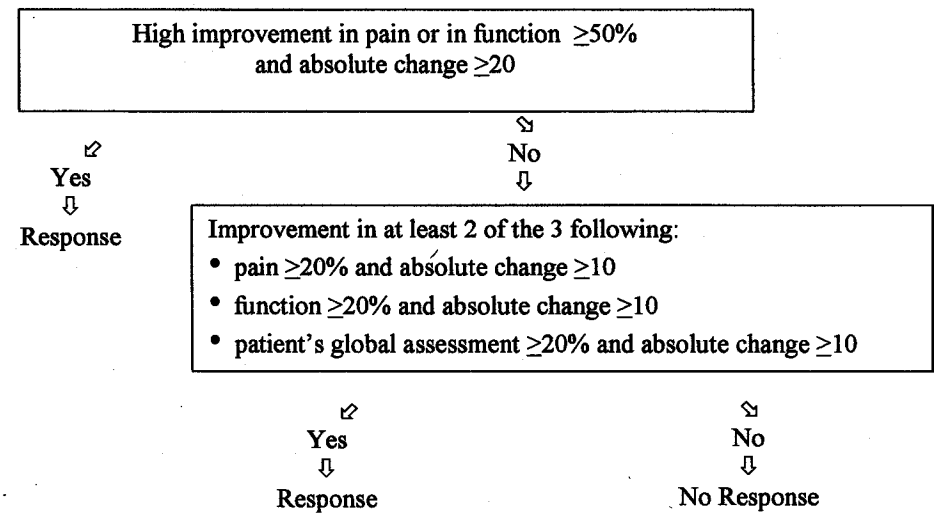


Figure 4. OMERACT-OARSI set of responder criteria.

to fulfill the set of criteria. The performance of such a set of criteria was observed in symptomatic OApatients for whom the level of symptoms required at entry was quite high. Further studies are required to evaluate performance of these criteria in other sets of patients, in particular, those with lower symptomatic activity at entry, but who are also candidates for joint structure preserving OA clinical trials.

ACKNOWLEDGMENT

The authors thank the following companies and their representatives for providing their database: Boehringer-Ingelheim France (Laurence Salin); Fidia Farmaceutici SpA (Simonetta Piva); Merck (James A. Bolognese, Elliot W. Ehrich, Ken E. Truitt); NEGMA-LERADS (Alain Taccoen, Patrick Cohen); Novartis (Ornella Della Casa Alberighi, Alan Moore); Pharmacia (Carl Wallemark, Jim Lefkowith); and Rotta Research Laboratorium (Lucio C. Rovati, Giampaolo Giacovelli).

REFERENCES

1. Altman RD, Hochberg MC. Degenerative joint disease. Clin Rheum Dis 1983;9:681-93.
2. Bellamy N, Kirwan J, Altman R, et al. Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip and hand osteoarthritis. Results of consensus development at OMERACTIII. J Rheumatol 1997;24:799-802.
3. Altman R, Asch E, Bloch D, 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.
4. Hochberg MC, Altman RD, Brandt KD, Moskowitz RW, for the Task Force: Design and conduct of clinical trials in osteoarthritis: Preliminary recommendations from a task force of the Osteoarthritis Research Society. J Rheumatol 1997;24:792-4.
5. Carlsson AM. Assessment of chronic pain. I. Aspects of the reliability and validity of the visual analogue scale. Pain 1983;16:87-101.
6. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol 1988;15:1833-40.
7. Lequesne MG, Mery C, Samson M, Gerard P. Indexes of severity for osteoarthritis of the hip and knee. Validation — value in comparison with other assessment tests. Scand J Rheumatol 1987;Suppl 65:85-9.
8. 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.
9. Dougados M, Leclaire P, van der Heijde D, Bloch DA, Bellamy N, Altman RD. 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.
10. O'Brien PC. Procedures for comparing samples with multiple endpoints. Biometrics 1984;40:1079-87.
11. Tilley BC, Pillemer SR, Heyse SP, Li S, Clegg DO, Alarcon GS. Global statistical tests for comparing multiple outcomes in rheumatoid arthritis trials. MIRA Trial Group. Arthritis Rheum 1999;42:1879-88.
12. Bolognese JA, Ehrich EW, Schnitzer TJ. Precision of composite measures of osteoarthritis efficacy in comparison to that of individual endpoints. J Rheumatol 2001;28:2700-4.