OsteoArthritis and Cartilage (2004) **12,** 389–399

doi:10.1016/j.joca.2004.02.001

Osteoarthritis and Cartilage

International Cartilage Repair Society

OSTEOARTHRITIS RESEARCH SOCIETY INTERNATIONAL

OMERACT-OARSI Initiative: Osteoarthritis Research Society International set of responder criteria for osteoarthritis clinical trials revisited.

T. Pham MD†, D. van der Heijde MD, PhD‡, R. D. Altman MD§, J. J. Anderson MD¶, N. Bellamy MD, MSc, FACP, FRCP, FRCPC, FAFRM, FRACP‖, M. Hochberg MD, MPH††, L. Simon MD‡‡, V. Strand MD§§, T. Woodworth¶¶ and M. Dougados MD‖‖*

† *René Descartes University, Cochin Hospital, Paris, Aix-Marseille II University, Conception Hospital, AP-HP, Marseille, France*

‡ *Department of Rheumatology, University Hospital, Maastricht, The Netherlands*

§ *University of Miami, Department of Veterans Affairs Medical Center, Miami, FL, USA*

¶ *Clinical Epidemiology Research and Training Unit, Boston Medical Center, Boston, MA, USA*

‖ *Department of Medicine, University of Queensland, Royal Brisbane Hospital, Queensland, Australia*

†† *Division of Rheumatology and Clinical Immunology, Department of Medicine, University of Maryland School of Medicine, Baltimore, MD, USA*

‡‡ *Medicine Department, Beth Israel Deaconess Medical Center, Boston, MA, USA*

§§ *Biopharmaceutical Consultant, Portola Valley, CA, USA*

¶¶ *Pfizer Global Research and Development, New London, CT, USA*

‖‖ *René Descartes University, Cochin Hospital, Paris, France*[1]

Summary

Background: The OARSI Standing Committee for Clinical Trials Response Criteria Initiative had developed two sets of responder criteria to present the results of changes after treatment in three symptomatic domains (pain, function, and patient's global assessment) as a single variable for clinical trials (1). For each domain, a response was defined by both a relative and an absolute change, with different cut-offs with regard to the drug, the route of administration and the OA localization.

Objective: To propose a simplified set of responder criteria with a similar cut-off, whatever the drug, the route or the OA localization.

Methods: Data driven approach:

(1) Two databases were considered

- The 'elaboration' database with which the formal OARSI sets of responder criteria were elaborated and
- The 'revisit' database.

(2) Six different scenarios were evaluated:

- The two formal OARSI sets of criteria
- Four proposed scenarios of simplified sets of criteria

Data from clinical randomized blinded placebo controlled trials were used to evaluate the performances of the two formal scenarios with two different databases ('elaboration' *versus* 'revisit') and those of the four proposed simplified scenarios within the 'revisit' database. The placebo effect, active effect, treatment effect, and the required sample arm size to obtain the placebo effect and the active treatment effect observed were the performances evaluated for each of the six scenarios. *Experts' opinion approach*: Results were discussed among the participants of the OMERACT VI meeting, who voted to select the definite OMERACT-OARSI set of criteria (one of the six evaluated scenarios).

*For correspondence and request for reprint: Professor Maxime Dougados, Rheumatology B Department, Cochin Hospital, 27 rue du faubourg Saint Jacques, 75014 Paris, France. Tel.: +33-1-58-41-25-62; Fax: +33-1-43-54-92-56; E-mail: maxime.dougados@cch.ap-hop-paris.fr

[1] Name of department and institution to which the work should be attributed: René Descartes University, AP-HP, Cochin Hospital, Rheumatology B Department, 27 rue du faubourg Saint Jacques, 75014 Paris, France.

Received 17 September 2002; revision accepted 3 February 2003.

Results: Data driven approach: Fourteen trials totaling 1886 OA patients and fifteen studies involving 8164 OA patients were evaluated in the 'elaboration' and the 'revisit' databases respectively.

The variability of the performances observed in the 'revisit' database when using the different simplified scenarios was similar to that observed between the two databases ('elaboration' *versus* 'revisit') when using the formal scenarios. The treatment effect and the required sample arm size were similar for each set of criteria. *Experts' opinion approach:* According to the experts, these two previous performances were the most important of an optimal set of responder criteria. They chose the set of criteria considering both pain and function as evaluation domain and requiring an absolute change and a relative change from baseline to define a response, with similar cut-offs whatever the drug, the route of administration or the OA localization.

Conclusion: This data driven and experts' opinion approach is the basis for proposing an optimal simplified set of responder criteria for OA clinical trials. Other studies, using other sets of OA patients, are required in order to further validate this proposed OMERACT – OARSI set of criteria.

Key words: Osteoarthritis, Outcomes, Clinical Trials Response Criteria Initiative.

Introduction

The Osteoarthritis Research Society International (OARSI) Standing Committee for Clinical Trials Response Criteria Initiative and the Outcome Measures in Rheumatology (OMERACT) committee, in concert with the international rheumatology community, has led to the development of a uniform core set of outcome measures for osteoarthritis (OA)[1–4]. One of the objectives was to propose a set of criteria for measurement based on multiple domains to present the results of changes after treatment in symptomatic parameters as a single variable for clinical trials. The symptomatic variables selected by both the OMERACT and OARSI societies were: pain, functional impairment and patient's global assessment.

Based on data from clinical trials, two sets of responder criteria (formal OARSI criteria) that can categorize an individual's response to treatment in a clinical trial have been developed[5] (Fig. 1).

The main characteristics of the proposed sets of criteria were the following:

- They covered three domains: pain, function and patient's global assessment.
- For each of these domains, a response was defined by both a relative and an absolute change.
- The cut-offs that defined a relevant change differed with regard to:
  - OA localization (*e.g.* hip *vs* knee),
  - evaluated study drug (*e.g.* NSAIDs *vs* specific anti-OA drug),
  - route of administration (*e.g.* per os *vs* intra-articular),
  - specific domain (pain, function, patient's global assessment).

The choice of the different cut-offs for the formal OARSI set of criteria was based on statistical analysis for optimization of the discriminant capacity. The preliminary attempts at uniform cut-off of all subsets showed a lesser placebo and active treatment effect of the set of criteria considered relevant by the members of the steering committee.

The main objective of this study was to evaluate the performances of the two previous formal OARSI sets of criteria and the performances of the modified ones, proposed by the scientific OMERACT committee. The aim of the proposed modifications was to simplify the presentation of the set of criteria, by evaluating different scenarios whatever the OA localization, whatever the evaluated drug, whatever the route of administration, and with similar cut-offs for the different domains.

Methods

PROPOSED SET OF CRITERIA

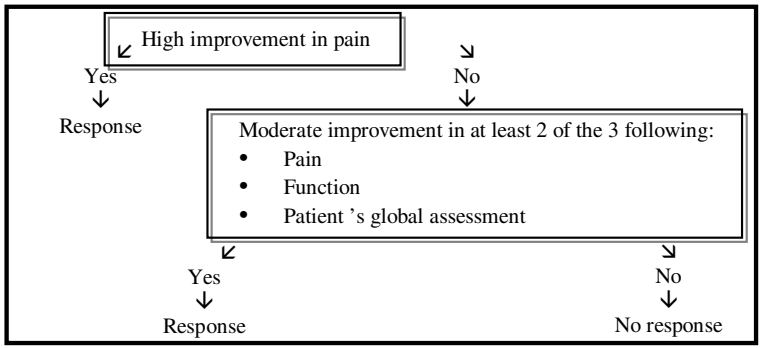
Six different scenarios were evaluated. The first two scenarios were the two propositions (A and B) of the formal OARSI set of criteria[5] (Fig. 1). The four other scenarios (scenarios C to F) were proposed by the OMERACT scientific committee. Their main characteristic was that they used a uniform cut-off whatever the OA localization, whatever the study drug and whatever the route of administration, unlike the formal OARSI set of criteria (Fig. 2). Scenarios A, C and E considered pain at the first responder step (high improvement), and scenarios B, D and F considered pain or function (Fig. 3). Scenarios C and D, as the formal OARSI set of criteria did, considered relative change (percentage of change during the study) and absolute change (absolute change during the study) in the variable to define a response, whereas scenarios E and F considered only relative change to define such response.

The study approach was both data driven and used an experts' opinion approach.

DATA DRIVEN APPROACH

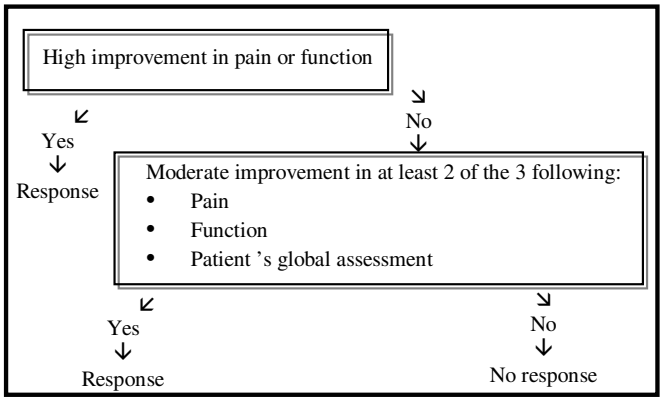
Two databases from clinical randomized placebo controlled trials were used:

- The initial one used to elaborate the formal set of criteria, known here as the 'elaboration database'.
- The second one is labeled the 'revisit database'. Drug companies who had conducted positive randomized placebo controlled trials in OA of a minimum 4-week duration were invited to revisit their database. The definition of 'positive' was based on a p value <0.05 for the *a priori* chosen primary criterion of the trial. Only the intention-to-treat analysis trials using the Last Observation Carried Forward technique were used. The participating drug companies were invited to provide anonymous information: OA localization, route of administration, characteristics of the study drug (analgesic, NSAID, Specific OA drug), study duration, number of patients in the placebo group and in the active treatment group, tools used to evaluate pain (*e.g.* pain VAS, Likert scale, WOMAC pain subscale), function (*e.g.* WOMAC function subscale) and patient's global assessment (*e.g.* VAS, Likert scale)[6], and time of collection of these different tools. Because of confidentiality, no demographic data, such as age, gender, body mass index, nor baseline values were asked to the drug companies. The drug was not identified by name, but only by class of agent (*e.g.* NSAIDs,



Subgroup	*High improvement in pain*				*Moderate improvement in* *Pain*		*Function*	
	Relative change*	Absolute change**	Relative change	Absolute change	Relative change	Absolute change	Relative change	Absolute change
Knee, oral NSAIDs	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, intra-articular specific drug	40	30	35	15	35	10	30	10

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



Subgroup	*High improvement in* *Pain*		*Function*		*Moderate improvement in* *Pain*		*Function*		*Global assessment*	
	Relative change*	Absolute change**	Relative change	Absolute change	Relative change	Absolute change	Relative change	Absolute change	Relative change	Absolute change
Hip, NSAIDs	50	30	50	20	25	15	20	10	20	10
Knee, oral NSAIDs	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, intra-articular specific drug	50	30	60	20	20	20	30	10	30	10

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

* Relative change: percentage of change during the study (final minus baseline over baseline × 100)
** Absolute change: absolute change during the study (final minus baseline on a 0-100 interval scale)

Fig. 1. OARSI Formal set of criteria: Scenarios A and B.

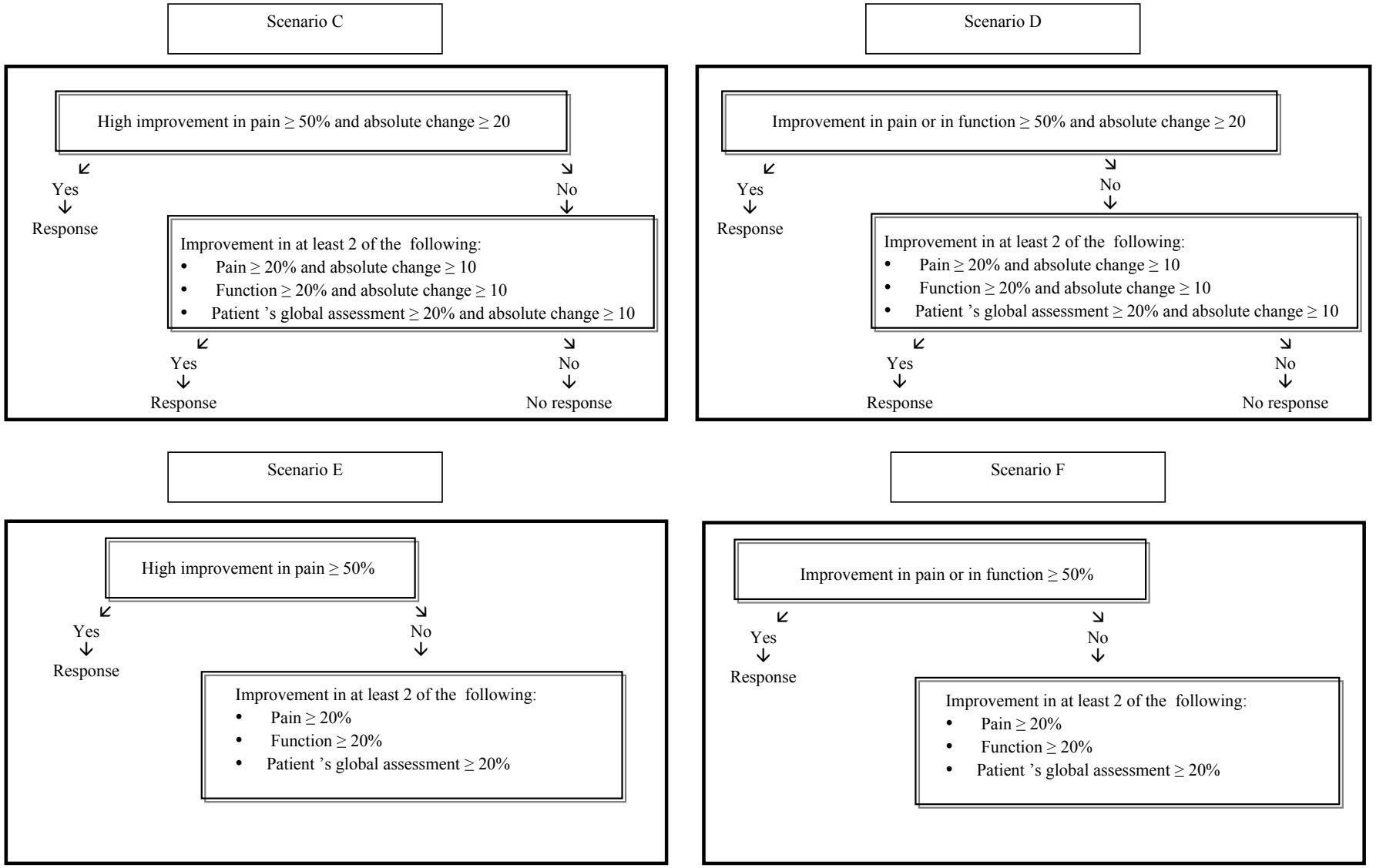


Fig. 2. Evaluated simplified sets of criteria: Scenarios C to F.

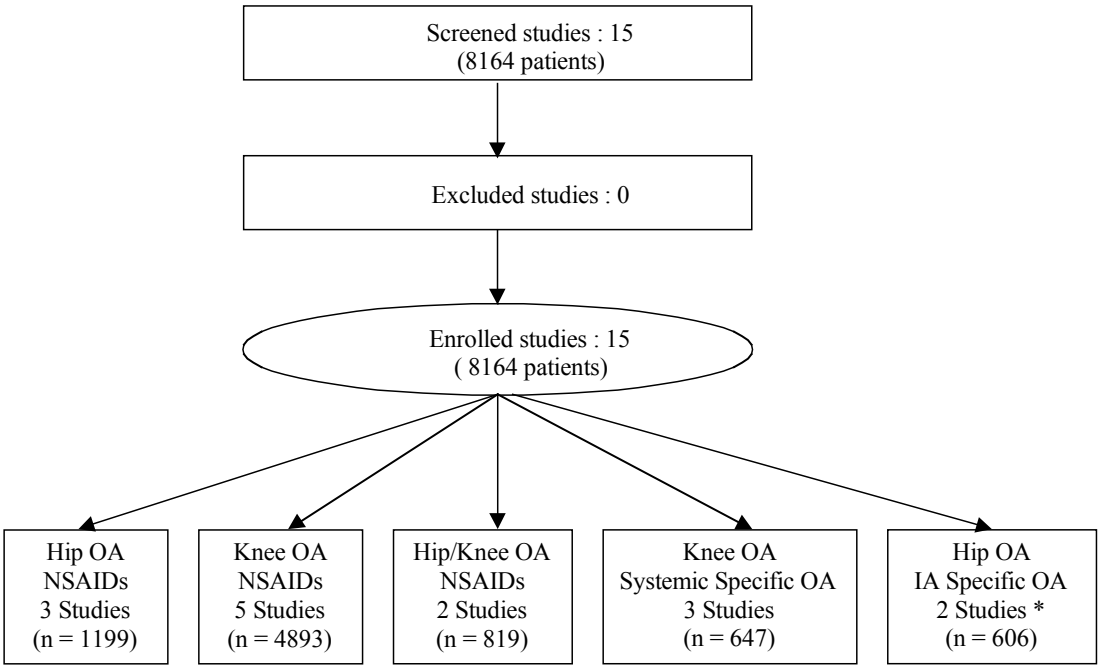


Fig. 3. Flow diagram for numbers of studies and patients.

systemic specific drug, intra-articular specific drug). For each trial and each scenario, a drug company provided the number of patients and the number of responders in each treatment arm. With this information, sensitivity (percentage of patients receiving an active drug labeled as responders according to the proposed set of criteria) and specificity (percentage of patients receiving the placebo treatment labeled as non-responder according to the proposed set of criteria) could be calculated for each trial and for each drug class and joint location of interest.

The first step of the data driven approach consisted in the evaluation of the performances of the two formal scenarios (scenario A and B), using the two databases, for each category of trial studied during the elaboration step (*i.e.* hip OA–NSAIDs trials, knee OA–NSAIDs trials, knee OA–systemic specific drug trials and knee OA–intra-articular specific drug trials). In other words, we compared the following performances in the 'elaboration' and in the 'revisit' database: placebo effect (percentage of responders in the placebo group), active effect (percentage of responders in the active treatment group), treatment effect (percentage of patients improved in the active treatment group minus the percentage of patients improved in the placebo group) and the sample arm size needed to obtain the observed placebo and active treatment effects (α=0.05 and β=0.20, two tailed test).

The second step consisted in the evaluation of the above performances between the six scenarios within the revisit database. For each drug and for each OA localization, the number of patients in the active treatment group and in the placebo group, the placebo effect, the active effect, the treatment effect and the sample arm size needed to obtain the observed placebo effect and the active treatment effect were calculated. These evaluations were also calculated whatever the localization and/or whatever the treatment. Moreover, since criteria sets almost always performed optimistically well when evaluated with the same database which was used to hunt for 'optimum' scenario, we compared the performances of the scenarios C, D, E and F in the elaboration database to the performances of the scenarios A and B in the revisit database.

Experts' opinion approach: Based on the data observed and after discussion among the OMERACT VI meeting participants, a vote was conducted to select the definite OMERACT-OARSI set of criteria (one of the six evaluated scenarios).

Lastly, the sensitivity and the specificity of the selected scenario has been evaluated in the "elaboration" database (knee OA–NSAIDs trials, hip OA–NSAIDs trials). The sensitivity was defined by the percentage of NSAIDs-OA patients meeting the OMERACT-OARSI criteria. The 1-specificity was defined by the percentage of placebo-OA patients meeting the OMERACT-OARSI criteria.

Results

PATIENTS AND STUDIES

In the elaboration database, fourteen trials totaling 1886 patients were evaluated (see Ref. 5 for details). The majority of the information was on NSAIDs for knee and hip. In the revisit database, fifteen studies involving 8164 OA patients were screened (Fig. 3). None of the studies was excluded. A prospective randomized controlled study in which the control group was receiving the usual therapeutic care without true placebo (*vs.* an intra-articular OA drug) was included in the revisit database. There were no trials

Table I
Characteristics of the 15 studies included in the 'revisit' database according to agent class*

Characteristics		Drug class		
		NSAIDs	Systemic specific OA drug	Intra articular specific OA drug
Number of studies		10	3	2
Number of patients	Active drug group	5557	316	303
	Placebo or control group	1354	331	303
Study duration	(mean +/- sd; weeks)	9.3±3.8	105.3±87.7	33±9.9
Pain evaluation	WOMAC	40%	100%	50%
	VAS	60%	0	50%
	Others	0	0	0
Function evaluation	WOMAC	100%	100%	50%
	VAS	0	0	50%
	Other	0	0	0
Global assessment evaluation	Likert	40%	33%	0
	VAS	30%	66%	50%
	Other	30%		50%
Time of collection of the outcome variables	Final and baseline	100%	100%	100%
	Only final visit	0	0	0

*'Revisit' data base is the one that permitted to revisit the formal sets of responder criteria and to evaluate the simplified sets of responder criteria.
WOMAC: Western Ontario McMaster Universities Osteoarthritis index; VAS: Visual Analogic Scale.

available to examine analgesics in OA. The majority of the studies concerned NSAIDs in hip and knee OA (10 of 15).

The characteristics of study designs are summarized in Table I. The data concerned 647 patients in systemic specific OA drug trials, 606 patients in intra-articular (IA) specific OA drug trials and 6911 patients in NSAIDs trials. For NSAIDs studies, whatever the OA localization, 5557 patients received the active treatment, and 1354 the placebo. Two studies involving hip and/or knee OA without indication of the localization were included only in the "whatever the localization" calculation. To assess pain and functional disability, two tools were most often used: The visual analog scale (VAS) and the Western Ontario McMaster Universities Osteoarthritis (WOMAC) index. For global patient's assessment, the VAS and the Likert scale were mostly used.

The knee was the only OA localization of the five specific OA drug studies (systemic and intra-articular), while NSAIDs trials were conducted in both knee and hip OA.

DATA DRIVEN APPROACH RESULTS

Formal set of criteria performances: comparison between elaboration database and revisit database

Results concerning the placebo and the treatment effects are summarized in Table II. For both the propositions A and B, the variability in the placebo and the active treatment effects were quite high (from 4% to 21% in the placebo group and from 7% to 34% in the active treatment group). Based on the observed results (placebo effect and active treatment effect) in the elaboration database, the calculation of the sample size required in future NSAIDs trials in knee OA was 67 patients per arm with scenario A and 66 with scenario B.

Performances of the 6 scenarios in the revisit database according to drug class, route of administration and OA localization

The results of the evaluated performance for each scenario are summarized in Table III. The highest active treatment effect and placebo effect were observed when using scenario F, whatever the drug class and whatever the localization.

NSAIDs in Knee OA. The highest active treatment effect was observed when using scenario F (66.4%), and at variance, the lowest placebo effect was observed 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 sizes "required" in future NSAID knee trials using the "revisit" data were 99 patients per arm, scenario A and 105 per arm, scenario B. Using the simplified scenarios, the sample sizes "required" were 98 per arm, scenario C, 101 per arm, scenario D, 97 per arm, scenario E and 98 per arm, scenario F.

NSAIDs in Hip OA. The highest active treatment effect was observed when using scenario F (60.8%), and at variance the lowest placebo effect was observed when using scenario A (28.9%). As observed in knee OA, the treatment effect was similar whatever the scenario (24.7%, 26.5%, 25.9%, 25.7%, 25.3% and 25.3% for scenarios A, B, C, D, E and F respectively). The sample sizes 'required' in future NSAID hip trials using the 'revisit' data were 62 patients per arm, scenario A and 55 per arm, scenario B. Using the simplified scenarios, the sample sizes 'required' were 58 per arm, scenario C, 59 per arm, scenario D and 61 per arm, scenario E and scenario F.

Systemic Specific OA drug in Knee OA. The highest active treatment effect was observed when using scenario F (49.4%), and the lowest placebo effect was observed when using scenario B (29.0%). Scenarios A and B showed the highest treatment effect (6.9% and 6.8% respectively) and the lowest sample size "required" for future systemic specific OA drug trials in knee OA (743 and 745 patients per arm respectively, *versus* 1167, 1095, 4979 and 3824 patients per arm for scenarios C, D, E and F).

Intra-articular specific OA drug in Knee OA. The highest active treatment effect was observed when using scenario F (72.9%), and the lowest placebo effect was observed when using scenario B (34.6%). The highest treatment

Table II

Performances observed with the formal sets of criteria, propositions 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 two 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 OA drug	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 OA IA Specific OA drug	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 OA NSAIDs	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 OA NSAIDs	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 section 3 of the manuscript for detailed explanations.

**'Elaboration' database is the one that permitted to propose the formal sets of responder criteria (5); 'Revisit' database is the one that permitted to revisit the formal sets of responder criteria and to evaluate the simplified sets of responder criteria.

Table III

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, β=0.20, two-tailed, expected placebo effect=that observed with this database, expected active treatment effect=that observed with this database*

Localization Drug		Knee OA Systemic Specific OA Drug	Knee OA IA Specific OA Drug	Hip OA NSAIDs	Knee OA NSAIDs	Whatever the joint NSAIDs	Whatever the joint Whatever the systemic treatment	Whatever the joint Whatever the treatment
Scenario A	% improved in active group	38.0%**	58.4%	53.6%	59.3%	58.3%	57.2%	57.2%
	% improved in placebo group	31.1%**	35.3%	28.9%	39.5%	36.8%	35.7%	35.7%
	Treatment effect	6.9%	23.1%	24.7%	19.8%	21.5%	21.5%	21.5%
	Sample size	745	73	62	99	84	84	84
Scenario B	% improved in active group	35.8%	57.4%	58.0%	58.4%	58.7%	57.5%	57.5%
	% improved in placebo group	29.0%	34.6%	31.5%	39.1%	37.4%	35.8%	35.6%
	Treatment effect	6.8%	22.8%	26.5%	19.3%	21.3%	21.7%	21.9%
	Sample size	743	74	55	105	86	82	81
Scenario C	% improved in active group	43.7%	70.3%	60.3%	65.1%	64.7%	63.6%	63.9%
	% improved in placebo group	38.0%	42.9%	34.4%	45.3%	43.3%	42.3%	42.2%
	Treatment effect	5.7%	27.4%	25.9%	19.8%	21.4%	21.3%	21.7%
	Sample size	1167	51	58	98	84	86	82
Scenario D	% improved in active group	44.6%	70.6%	60.5%	65.4%	65.0%	63.9%	64.2%
	% improved in placebo group	38.7%	43.6%	34.8%	45.9%	43.9%	42.8%	42.9%
	Treatment effect	5.9%	27.0%	25.7%	19.5%	21.1%	21.1%	21.3%
	Sample size	1095	52	59	101	87	87	85
Scenario E	% improved in active group	47.8%	72.6%	60.5%	65.8%	65.3%	64.3%	64.7%
	% improved in placebo group	45.0%	44.5%	35.2%	45.9%	44.0%	44.2%	44.3%
	Treatment effect	2.8%	28.1%	25.3%	19.9%	21.3%	20.1%	20.4%
	Sample size	4979	48	61	97	85	96	93
Scenario F	% improved in active group	49.4%	72.9%	60.8%	66.4%	66.0%	65.1%	65.5%
	% improved in placebo group	46.2%	45.2%	35.5%	46.6%	44.5%	44.8%	44.9%
	Treatment effect	3.2%	27.7%	25.3%	19.8%	21.5%	20.3%	20.6%
	Sample size	3824	49	61	98	83	94	91

*'Revisit' database is the one that permitted 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 was observed when using scenario E (28.1%). The lowest sample size 'required' for future intra- articular specific OA drug trials in knee OA were observed when using the simplified scenarios (51, 52, 48 and 49 patients per arm for scenarios C, D, E and F respectively, *versus* 73 and 74 patients per arm for scenarios A and B).

Performances of the six scenarios in the revisit database whatever the drug class, the route of administration or the localization of OA

NSAIDs whatever the OA localization. The highest active treatment effect was observed when using scenario F (66.0%), and at variance the lowest placebo effect was observed when using scenario A (36.8%). The treatment effect was similar whatever the scenario (21.5%, 21.3%, 21.4%, 21.1%, 21.3% and 21.5% for scenarios A, B, C, D, E and F respectively). The sample size "required" for future NSAIDs trials in OA was also similar whatever the scenario (84, 86, 84, 87, 85 and 83 patients per arm for scenarios A, B, C, D, E and F respectively).

Whatever the systemic drug (i.e. systemic specific OA drugs and NSAIDs) and whatever the localization. The highest active treatment effect was observed when using scenario F (65.1%), and at variance the lowest placebo effect was observed when using scenario B (21.7%). The treatment effect was similar whatever the scenario (21.5%, 21.7%, 21.3%, 21.1%, 20.1% and 20.3% for scenarios A, B, C, D, E and F respectively). The "required" sample size was also similar whatever the scenario (84, 82, 86, 87, 96 and 94 patients per arm for scenarios A, B, C, D, E and F respectively).

Whatever the drug and whatever the localization. The highest active treatment effect was observed when using scenario F (65.5%), and at variance the lowest placebo effect was observed when using 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" for future trials in OA was 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).

EXPERTS' OPINION APPROACH RESULTS

Based on the observed results, it was considered that the data driven approach did not permit to select a specific set of criteria. However, at least two of these performances (treatment effect and required sample size) were similar whatever the scenario (A to F). These results were presented to the participants of the Osteoarthritis session of the OMERACT VI conference (Brisbane 2002). After discussion and voting, it appears that:

- The treatment effect and the required sample size were the two major characteristics to take into account in the choice of an optimal set of criteria to be used for clinical trials.
- Two other characteristics were also considered as important:

1) The definition of an improvement based not only on a relative change but also on an absolute change (scenarios A, C and E *versus* scenarios B, D and F)
2) The simplicity of the presentation: same cut-offs, set of responder criteria whatever the localization, the study drug and the route of administration (scenario A, B *versus* C, D, E and F).

Based on this preliminary discussion between experts and after a voting session, scenario D was selected (Fig. 4). It is now labeled the 'OMERACT-OARSI' set of responder criteria.

EVALUATION OF THE DIFFERENT SETS OF CRITERIA

Table IV summarizes the results of the procedure permitting the evaluation of the different scenarios. This table shows that the treatment effect was similar whatever the evaluated scenario, but for hip OA, both the sensitivity and the specificity (active treatment effect and placebo treatment effect) were higher for the scenario D.

Discussion

This study, which combined the efforts of academic researchers, representatives of the pharmaceutical industry and representatives of health agency, proposes a simplified set of responder criteria for clinical trials in OA by simplifying the initial OARSI set of criteria using a data driven and experts' opinion approach. Limitations of this study include (i) the absence of analgesics trials in our analysis of the improvement between active drug-treated group and placebo-treated group; (ii) Available trials concerned only knee or hip OA and no other OA localization; (iii) The collected data concerned only the core set of criteria. Drug companies provided for each trial, the percentage of responders in the active treatment group and the percentage of non-responders in the placebo group, according to each scenario. We did not have access to the individual data, neither to the percentage of responders for each domain separately (pain, function, global patient's assessment). This lack of data did not allow us to estimate if the core set of criteria was less powerful than each domain treated separately, as has been done for rheumatoid arthritis[7]; (iv) The cut-offs of the simplified scenarios were inspired by the formal ones. However, more specific cut-offs could not be estimated due to the lack of individual data for the 8164 OA patients.

We observed considerable variability in the results with regard to the study population (elaboration *versus* revisit database) within the formal sets of criteria. This variability could be attributed to a variability between the patients included in the two databases. However, in both of them, most of the trials have been conducted in multicenter international trials following a very similar approach concerning the inclusion and exclusion criteria (phase II and phase III trials).

In the elaboration phase of the formal OARSI set of criteria, the loss of sensitivity and specificity using identical cut-offs, whatever the localization and the study drug, did not allow to propose a simple set of criteria (similar cut-off whatever the OA localization and the study drug). The variability of the performances of these formal sets of criteria between the two databases was in contradiction with the results obtained in the elaboration phase and prompted us to further evaluate a simplification of the set of responder criteria.

The data driven conclusions are that, whatever the OA localization, the study drug or the route of administration, formal scenarios A and B had the lowest placebo effect, and scenario F had the highest active treatment effect. In

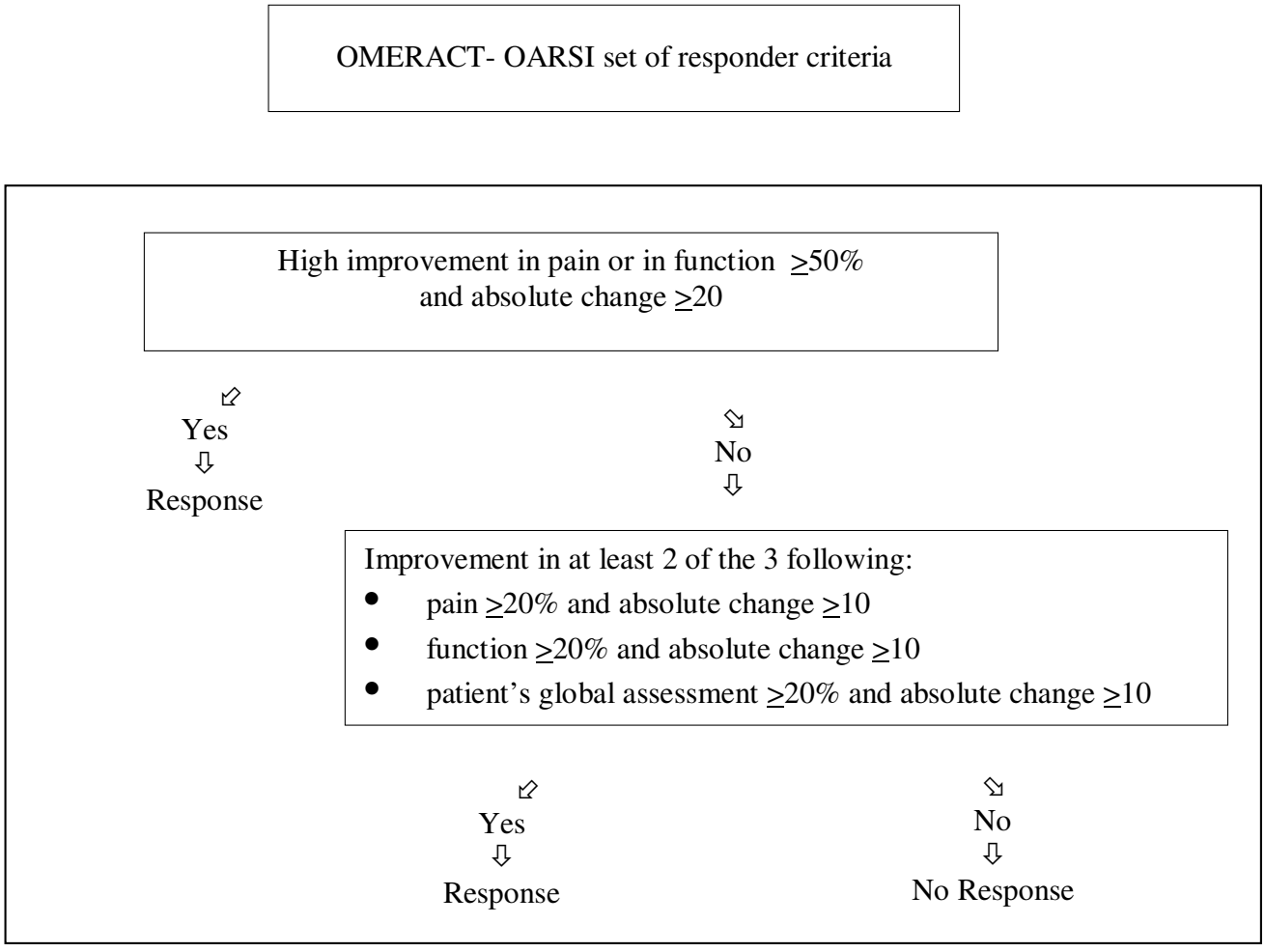


Fig. 4. OMERACT–OARSI Set of responder criteria.

Table IV
*Percentage of patients responding when the OMERACT-OARSI and formal proposition A and B criteria sets are applied to a validation data set**

Validation database	Criteria set	Knee NSAID trial		Hip NSAID trial	
		Sensitivity	Specificity	Sensitivity	Specificity
Elaboration	OMERACT-OARSI	59%	40%	72%	44%
Revisit	OARSI-Proposition A	59%	39%	54%	29%
Revisit	OARSI-Proposition B	58%	39%	58%	26%

*Sample size required per arm alpha=0.05, beta=0.20, two-tailed.Sensitivity=% responders on active drug (NSAIDs). 1-Specificity=% responders on placebo.

contrast, the treatment effect and the required sample size were quite similar whatever the scenario, ranging from 20.4% to 21.9% and from 81 to 93 patients per arm respectively.

Although the data driven approach did not allow to select any particular scenario, the simplification of the set of criteria did not result in a loss of relevant performances. Indeed, a higher active treatment effect and a higher placebo effect were observed when using the simplified scenarios in both databases. Conversely, the treatment effect and the sample size required to obtain the observed placebo and active effects were similar whatever the scenario (whether formal or simplified) in the revisit database.

According to the experts, these two performances were the most important for an optimal set of responder criteria. Although all the evaluated scenarios provided similar results for these performances, the experts' choice was scenario D (Fig. 4), which confirms the importance of:

1) A format that requires both an absolute change and a relative change.
2) A format that considers both pain and function as important domains; in certain studies, however, changes in functional disability are at least as important as changes in pain.

The observed treatment effect whatever the drug and whatever the treatment when using scenario D is 21.3%. This result is close to what is expected in OA, i.e., 20-30%[8,9].

The required sample size with scenario D whatever the drug and whatever the treatment is 85 patients per arm. This is similar to the sample size required when using the previous formal set of criteria.

In conclusion, we propose a simplified definition for symptomatic improvement in osteoarthritis. This set of criteria, approved both by the OARSI and the OMERACT committees, is at least as powerful as the previous OARSI formal set of criteria and its simplification will probably enhance its use in future OA trials.

Other studies are required on order to further validate this proposed OMERACT-OARSI set of criteria in other sets of patients suffering from osteoarthritis of different localizations and treated differently, e.g., with analgesics or non-pharmacological therapies.

Acknowledgements

The authors wish to thank the drug companies and their representatives for their active collaboration: 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),* Rotta Research Laboratorium (*Lucio C. Rovati, Giampaolo Giacovelli).*

References

1. Altman R, Brandt T, Hochberg M, Moskovitz R. 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.
2. Bellamy N, Kirwan J, Altman R, Boers M, Brandt K, Brooks P, *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 OMERACT III. J Rheumatol 1997;24:799–802.
3. Lequesne M, Brandt K, Bellamy N, Moskovitz R, Menkes CJ, Pelletier JP, *et al.* Guidelines for testing slow-acting drugs in osteoarthritis. J Rheumatol 1994;41(Suppl):65–71 discussion 72–3. (Published erratum appears in J Rheumatol Suppl 1994 Dec; 41: 2395.
4. Group for the respect of ethics and excellence in science (GREES): osteoarthritis section. Recommendations for the registration of drugs used in the treatment of osteoarthritis. Ann Rheum Dis 1996;55(8): 552–7.
5. 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(6): 395–403.
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(12):1833–40.
7. Felson DT, Anderson JJ, Boers M, Bombardier C, Furst D, Goldsmith C, *et al.* American College of Rheumatology. Preliminary definition of improvement in rheumatoid arthritis. Arthritis Rheum 1995;38(6): 727–35.
8. 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(8):1039–49.
9. Bolnot-Delmas D, Buch JP, Zeidler H, Dougados M. Ro 15-8081 in osteoarthritis of hip and knee: a double-blind placebo-controlled multicentre dose-ranging study on analgesia. Pain 1996;64(1):99–105.