

Recommendations for a Core Set of Outcome Measures for Future Phase III Clinical Trials in Knee, Hip, and Hand Osteoarthritis. Consensus Development at OMERACT III

NICHOLAS BELLAMY, JOHN KIRWAN, MAARTEN BOERS, PETER BROOKS, VIBEKE STRAND, PETER TUGWELL, ROY ALTMAN, KENNETH BRANDT, MAXIME DOUGADOS, and MICHEL LEQUESNE

ABSTRACT. Significant progress has been made in outcome measurement procedures for osteoarthritis (OA) clinical trials, and guidelines have been established by the US Food and Drug Administration, European League Against Rheumatism, the World Health Organization/International League of Associations for Rheumatology, and the Group for the Respect of Ethics and Excellence in Science. However, there remains a need for further international harmonization of measurement procedures used to establish beneficial effects in Phase III clinical trials. A key objective of the OMERACT III conference was to establish a core set of outcome measures for future phase III clinical trials. During the conference, using a combination of discussion and polling procedures, a consensus was reached by at least 90% of participants that the following 4 domains should be evaluated in future phase III trials of knee, hip, and hand OA: pain, physical function, patient global assessment, and, for studies of one year or longer, joint imaging (using standardized methods for taking and rating radiographs, or any demonstrably superior imaging technique). These evidence based preferences, achieved with a high degree of consensus, establish an international standard for future phase III trials and will also facilitate metaanalysis and Cochrane Collaborative Project goals. (*J Rheumatol* 1997;**24**:799–802)

Key Indexing Terms:

OSTEOARTHRITIS ENDPOINTS CORE SET OUTCOME MEASURES

Outcome measurement in clinical trials requires the use of valid, reliable, and responsive measurement procedures that adequately capture important aspects of the condition. In recognition of this requirement, a number of individuals and groups have published lists of recommended outcome measures¹⁻⁵. In particular, the US Food and Drug Administration, European League Against Rheumatism, World Health Organization/International League of Associations for Rheumatology, and the Group for Respect of Ethics and Excellence in Science have published guidelines which in part specify domains and in part recognize

actual measurement techniques or instruments. While not in complete agreement, the existing guidelines nevertheless share several important elements, namely, the measurement of pain, walk time, patient global assessment, and physician global assessment.

To build on experience and current preference but not exclude other measures of potential importance in future trials, a process was followed that had 4 basic elements: (1) provision of information from the literature; (2) lectures followed by discussion periods; (3) breakout groups; (4) polling procedures.

THE PROCESS

Prior to OMERACT III, participants were asked to complete an initial questionnaire to identify candidate variables. From return questionnaires, a second questionnaire was then constructed incorporating additional suggestions. The questionnaire was extensive and identified 4 site specific forms of osteoarthritis (OA) (knee, hip, hand, and generalized), 2 types of studies (symptom modifying OA drugs and structure modifying OA drugs), 3 levels of measurement (clinical, imaging, and biologic markers) and various domains and measurement techniques. Participants were asked to rank in order of importance their preferences for outcome measurement for each clinical situation and drug class. This proved excessively demanding and only 15 questionnaires

From London Health Sciences Centre, Victoria Campus, London, Canada; Bristol Royal Infirmary, Bristol, UK; University Hospital Maastricht, Maastricht, The Netherlands; St. Vincent's Hospital, Darlinghurst, Australia; Stanford University, San Francisco, USA; University of Ottawa, Ottawa, Canada; Miami School of Medicine, Miami, USA; Indiana University School of Medicine, Indianapolis, USA; Hôpital Cochin and Collège de Médecine des Hôpitaux de Paris, Paris, France.

N. Bellamy, MD, Victoria Hospital; J. Kirwan, MD, Royal Bristol Infirmary; M. Boers, University Hospital Maastricht; P. Brooks, MD, St. Vincent's Hospital; V. Strand, Stanford University; P. Tugwell, MD, University of Ottawa; R. Altman, MD, Miami School of Medicine; K. Brandt, Indiana University School of Medicine; M. Dougados, MD, Hôpital Cochin; M. Lequesne, MD, Collège de Médecine des Hôpitaux de Paris.

Address reprint requests to Dr. N. Bellamy, Ste. 303 Colborne, Bldg., London Health Sciences Centre, 375 South Street, London, ON N6A 4G5, Canada.

were returned. Prior to OMERACT III each participant also received position papers that outlined the dimensionality of the measurement problem and provided up-to-date information in areas of clinical, imaging, and biologic markers.

During OMERACT III, participants attended presentations addressing the different measurement areas and, where available, data were presented on the clinimetric properties of different instruments and comparisons of measurement techniques. Time was allowed during question period for clarification and for alternative viewpoints. Participants then completed an exercise in which they were asked to assign 100 points to reflect their measurement preferences in each of 4 types of OA trials (knee, hip, hand, generalized). Participants next separated into 3 clinical and one combined imaging/biologic markers group. Feedback was available from the voting profiles within each of the breakout groups. The breakout groups provided an opportunity to discuss contentious issues more fully and bring back recommendations to the group as a whole. Following these deliberations as well as other informal discussions, a final questionnaire was designed to allow participants to vote for inclusion of domains in a core set and to express use preferences for types of instruments. However, questions regarding specific instruments, while permitting flexibility, were not generated from prior voting procedures and a decision was made not to include recommendations regarding specific instruments for research applications.

THE CONSENSUS

Participants were provided opportunity to recommend a measure for inclusion in (a) the core set (i.e., mandatory in future Phase III clinical trials in knee, hip, and hand OA studies); (b) the research agenda (i.e., worthy of further formal evaluation and possible future inclusion in the core set); or (c) inclusion in neither the core set nor the research agenda. The summary results are shown in Table 1.

After presentation of these data a number of issues were raised.

1. Whether generalized OA was a distinct and definable entity for clinical trials purposes. (Resolution — to exclude further consideration of generalized OA.)
2. Whether the rate of onset of therapeutic effect (fast versus slow) determined the need for different types of clinical measures. (Resolution — time of onset determines when to measure rather than what to measure.)
3. Whether different measures were required for an analgesic study versus a nonsteroidal antiinflammatory drug (NSAID) study. (Resolution — the domains are the same but the measurement techniques might vary.)
4. Whether clinical measures should be different for system modifying versus structure modifying OA drug studies. (Resolution — the clinical core domains are the same.)
5. It was assumed that biologic markers would be important in the future, but confirmatory evidence is lacking for the evaluative and predictive value of any single market.
6. It was acknowledged that data existed on the value of measures of health related quality of life (generic and utility measures), but that no one measure had yet been identified as superior to all others for clinical trial purposes. The importance of such measures in health related quality of life determination, cross study and cross disease comparisons, and in pharmacoeconomic comparisons was generally acknowledged. As a result, while not in the core set, it was decided to strongly recommend the incorporation of health related quality of life measures in future Phase III trials of at least 6 months' duration. Over the next 3 to 5 years it should be possible to evaluate the role of such measures in clinical trials.
7. It was emphasized that no measure was excluded from use in future clinical trials by decisions made at OMERACT III. Indeed, in some studies the primary outcome might not be one cited in the core set (e.g., the effect of a future drug on time to surgery). However, such studies would be

Table 1. Preferences for core set of efficacy domains in future Phase III hip, knee, and hand OA trials.

Domain	In Core (% Voting Yes)	In Research Agenda (% Voting Yes)	In Neither (% Voting Yes)	Number Voting
Pain	100	0	0	75
Physical function	97	1	1	76
Imaging* (in studies of 1 yr or longer)	92	7	1	76
Patient global assessment	91	1	1	75
Physician global assessment	52	21	27	73
Generic quality of life/utility	36	58	6	69
Stiffness	14	61	25	72
Other**	13	69	19	16
Inflammation	8	70	22	74

* Standardized techniques for taking and scoring radiographs or demonstrably superior imaging techniques.

** Includes tenderness, performance based measures, time to surgery, number of flares, biologic markers.

required to also include assessments of domains cited in the core set in the measurement battery.

8. There was debate whether stiffness should be incorporated, whether pain and stiffness were part of the same domain, whether patients understood the concept of stiffness, and whether current techniques accurately assessed it. (Resolution — when stiffness is to be assessed in hip and/or knee studies it should be measured using the WOMAC or Algofunctional Severity Indices.)

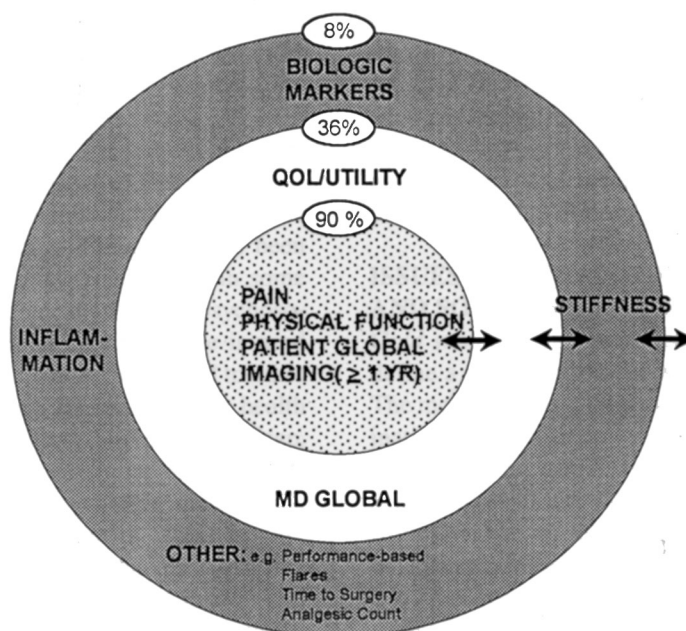
9. There was debate on the value of physician global assessment in OA trials (as there had been at OMERACT I regarding its use in rheumatoid arthritis trials). Only 52% of participants felt it should be included in the core set for OA and as a result it was not included. It was acknowledged, however, that it was important to about half the participants and its continued use was acceptable.

In drawing up the core set, 3 assumptions were proposed; (1) to be included there needed to be evidence for reliability, validity, and responsiveness; (2) it was not necessary to specify exact instruments, but only to agree on the major domains to be included; (3) there is a difference between consensus and unanimity. However, a 51/49% split seemed

insufficient, since 49% of participants would be in disagreement. Similarly, a 60/40% split would not be decisive. Common sense suggests if 90% or more participants agreed on a core set, one could claim a consensus, albeit without unanimity. As a result the core set recommended by OMERACT III was based on a consensus of $\geq 90\%$ and included the following measures:

- Pain
- Physical function
- Patient global assessment
- Imaging in studies ≥ 1 year (As an efficacy measure in structure modifying OA drug studies, but also as a safety measure in pure system modifying OA drug studies of ≥ 1 year duration.)

These are illustrated in Figure 1, in which the inner core defines the core set of OA. The middle core identifies health related quality of life measures (optional, but strongly recommended) and physician global assessment (optional, depending on perceived importance to the investigator). The outer core contains measures of stiffness (by WOMAC and Algofunctional Severity Indices), biologic markers, measures of inflammation, and other assessments (e.g., performance based measures, flares, time to surgery, anal-



% voting for inclusion in core set	Placement	Consequence
$\geq 90\%$	INNER CORE →	"CORE SET"
$\geq 36\% - < 90\%$	MIDDLE CORE →	QOL/UTILITY (Strongly Recommended)
$8\% - < 36\%$	OUTER CORE →	OPTIONAL

Figure 1. Osteoarthritis core concept.

gesic consumption), all of which are optional measures. This concept places highly patient relevant measures at the center, while measures less relevant to patients are at the periphery. It should be noted that only domains cited in the inner core (i.e., core set) will be obligatory in outcome measurement in future Phase III trials. Any instrument used should be of adequate reliability, validity, and responsiveness. For imaging, the preferred technique currently is radiographic and requires standardized methods for both taking and scoring films. The term imaging was selected specifically to allow for future developments of technically superior methods.

CONCLUSION

These evidence based preferences were achieved through a high degree of consensus. They allow international harmonization of outcome measurement procedures in OA clinical trials. However, they also offer 4 additional advantages: (1) they do not exclude other measures being used in addition to the core set; (2) they are flexible and allow over time for the inward and outward migration of measures as developments occur in clinical, imaging, and molecular disciplines; (3) they create a foundation on which other organizations and consensus conferences can build, particularly with respect to the specification of exact instruments for use in specific situations; and (4) they will facilitate metaanalyses and Cochrane Collaborative Project goals⁶.

In summary, participants at OMERACT III agreed ($\geq 90\%$) on a core set of 4 domains for outcome measurement in future Phase III clinical trials of hip, knee, and hand OA. The 4 domains identified were pain, physical function, patient global assessment, and, for studies of at least one year, joint imaging.

REFERENCES

1. Guidelines for the Clinical Evaluation of Anti-inflammatory and Anti-rheumatic Drugs (Adults and Children). US Department of Health and Human Services. Public Health Services, Food and Drug Administration. Rockville, MD: April 1988:12-5.
2. Guidelines for the Clinical Investigations of Drugs Used in Rheumatic Diseases. European Drug Guidelines, Series 5, World Health Organization, Regional Office for Europe, Copenhagen, European League Against Rheumatism. March 1985:21-4.
3. Lequesne M, Brandt K, Bellamy N, *et al*: Guidelines for testing slow acting drugs in osteoarthritis. *J Rheumatol* 1994;(suppl 41)21:65-73.
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;55:552-7.
5. Bellamy N: Osteoarthritis. In: *Musculoskeletal Clinical Metrology*. Dordrecht: Kluwer Academic Publishers, 1993:169-76.
6. Cochrane AL: Effectiveness and Efficiency. Random Reflections on Health Services. London: Nuffield Provincial Hospitals Trust, 1972. (reprinted) London: *BMJ*, 1989.