
Guidance for Industry

Clinical Development Programs for Drugs, Devices, and Biological Products Intended for the Treatment of Osteoarthritis (OA)

DRAFT GUIDANCE

This guidance document is being distributed for comment purposes only.

Comments and suggestions regarding this draft document should be submitted within 60 days of publication of the *Federal Register* notice announcing the availability of the draft guidance. Submit comments to Dockets Management Branch (HFA-305), Food and Drug Administration, 5630 Fishers Lane, rm. 1061, Rockville, MD 20852. All comments should be identified with the docket number listed in the notice of availability that publishes in the *Federal Register*.

For further information on this draft guidance, contact Sandra Cook, Center for Drug Evaluation and Research (HFD-550), 9201 Corporate Blvd., Rockville, MD 20850, 301-827-2090.

**U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)
Center for Devices and Radiological Health (CDRH)
July 1999**

Guidance for Industry

Clinical Development Programs for Drugs, Devices, and Biological Products Intended for the Treatment of Osteoarthritis (OA)

Copies of this Guidance are available from:

*Office of Training and Communications
Division of Communications Management
Drug Information Branch, HFD-210
Center for Drug Evaluation and Research
Food and Drug Administration
5600 Fishers Lane, Rockville, MD 20857
(Phone 301-827-4573)*

Internet: <http://www.fda.gov/cder/guidance/index.htm>.

or

*Office of Communication, Training and
Manufacturers Assistance, HFM-40
Center for Biologics Evaluation and Research
Food and Drug Administration
1401 Rockville Pike, Rockville, MD 20852-1448
Internet: <http://www.fda.gov/cber/guidelines.htm>.
Fax: 1-888-CBERFAX or 301-827-3844*

Mail: the Voice Information System at 800-835-4709 or 301-827-1800.

or

*The Division of Small Manufacturers Assistance (DSMA), HFZ-220
Center for Devices and Radiological Health
Food and Drug Administration
1350 Piccard Drive, Rockville, MD 20850
800-638-2041 or 301-443-6597
Internet: DSMA@CDRH.FDA.GOV
Fax: 1-301-443-8818*

Facts-On-Demand (FAX) at 800-899-0381 or 301-827-0111

**U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)
Center for Devices and Radiological Health (CDRH)
July 1999**

Table of Contents

I.	INTRODUCTION.....	1
II.	USE OF PRECLINICAL MODELS	2
III.	PRODUCT DEVELOPMENT.....	2
IV.	OSTEOARTHRITIS MEASUREMENTS	3
	A. Pain and Function	3
	B. Structure	4
V.	OSTEOARTHRITIS CLAIMS	5
	A. Treatment of Symptoms: Pain and Function	5
	B. Delay in Structural Progression	6
	C. Prevention of OA	7
VI.	TRIAL DESIGNS AND ANALYSES	7
VII.	ASSEMBLING THE EVIDENCE	8
VIII.	OVERALL RISK-BENEFIT ASSESSMENT	8

Draft — Not for Implementation
Guidance for Industry¹

**Clinical Development Programs for
Drugs, Devices, and Biological Products
Intended for the Treatment of Osteoarthritis (OA)**

I. INTRODUCTION

This draft guidance is intended to assist sponsors who are developing drugs, devices, or biological products to treat human osteoarthritis (OA). The guidance discusses a number of issues that should be considered during development, such as the utility of animal models and the measurement of improvement in human OA. The guidance also discusses the types of label claims that can be considered for such products and provides guidance on the clinical development programs to support these claims. The central purpose of label claims is to inform prescribers and patients about the documented benefits of a product.

Because OA is a disease with a complex pathophysiology, multiple clinical outcomes can be envisioned. This guidance discusses a number of potential claims, including an *improvement in symptoms* claim, and a *delay in structural progression* claim. The guidance also proposes, in principal, a *prevention of OA* claim. The purpose of this draft is to continue public discussion of the concepts underlying these claims and the nature of the evidence to support them.

Many of the topics in OA disease assessment, clinical trial design, and analysis have parallels in rheumatoid arthritis (RA) disease assessment. As a result, sponsors should also refer to the Agency's guidance for industry on *Clinical Development Programs for Drugs, Devices, and Biological Products for the Treatment of Rheumatoid Arthritis (RA)* (FDA 1999).

¹ This guidance has been prepared by the Rheumatology Working Group of the Medical Policy Coordinating Committee (MPCC) in the Center for Drug Evaluation and Research (CDER) at the Food and Drug Administration. The working group includes members from the Center for Biologics Evaluation and Research (CBER) and the Center for Devices and Radiological Health (CDRH). This guidance document represents the Agency's current thinking on osteoarthritis. It does not create or confer any rights for or on any person and does not operate to bind FDA or the public. An alternative approach may be used if such approach satisfies the requirements of the applicable statutes, regulations, or both.

II. USE OF PRECLINICAL MODELS

Human OA is a chronic disease of the joints causing pain and dysfunction; it is sometimes debilitating. OA is characterized by pain, biochemical and enzymatic changes, cartilage fragmentation and loss, osteophyte formation, and bony sclerosis. These symptoms, or processes, differ in their clinical effect, depending on the particular joint affected. Certain risk factors for OA have been identified, including trauma (singly or repetitive), anatomic/postural abnormalities including joint instability, and a putative genetic predisposition. These risk factors should be addressed, and preclinical models may be useful in gaining a better understanding of these factors. For example, an agent selectively targeting pain might also cause toxicity by a different mechanism. Using an animal model might enable the screening of analogs to maximize one effect while minimizing the other.

Although the same fundamental processes are probably involved in joint destruction in animal models of OA, there are important species differences in the relative contribution of inflammatory mediators. Therefore, the usefulness of an animal OA model will depend on the location and function of the involved joints, as well as on the animal's particular pathophysiology.

Compared to RA, few models of human OA are currently in use. Examples include the guinea pig spontaneous OA model and the Pond Nuki dog model. When evaluating the possible usefulness of an animal model, the following questions should be considered

1. How accurately does the model replicate human OA?
2. What are the structural determinants of pain and loss of function?
3. Do structural changes (identified with MRI, x-ray) correlate with clinical (pain, motion, weight distribution, gait) or biochemical (cartilage composition, enzymatic activity, pain mediators, receptor expression) markers?
4. Is the model useful for studying prophylactic strategies or for studying structural arrest or reversal?
5. Can the model be used to assess long-term toxicity?

Product developers can incorporate toxicity endpoints into a model, or design parallel assessments of activity and toxicity. Either strategy will enhance the usefulness of animal data in defining safety margins. Because the eventual utility of a product depends on the risk-benefit — particularly the long-term safety margin in older populations — toxicologic evaluation may help predict and characterize a product's safety profile.²

III. PRODUCT DEVELOPMENT

² More information about toxicologic evaluation of agents for chronic use can be found in standard FDA guidances.

Draft — Not for Implementation

Establishing the optimal clinical dose for products with delayed onset of effect is difficult and may call for extrapolating from biochemical or pharmacodynamic findings. This is particularly true for products aimed at retarding the progression of OA as measured by joint space narrowing (JSN), currently the best accepted marker for structure. This is because a product's effects on JSN and symptoms are presumed to be very slow in onset. In the future, MRI (magnetic resonance imaging) may enable dose-finding trials for products to treat JSN to be shortened. Findings from ongoing research in cartilage metabolism, bone density, structural assessments, and arthroscopic measurements (visual, biomechanical, histological), and their correlation with traditional measurements may facilitate product development. Future findings may also help avoid or at least better characterize unanticipated toxicities.

IV. OSTEOARTHRITIS MEASUREMENTS

Because current treatment in OA is fundamentally symptomatic and not thought to affect long-term outcomes, trial experience has been limited almost exclusively to the study of short-term hypotheses. Protocols enrolling patients with knee or hip OA (the so-called signal joints) have made measuring and interpreting treatment effects easier, and the development of specific OA measurements has paralleled, and in some ways guided, this signal-joint approach. However, exclusive focus on the signal-joint will miss what is happening at other OA sites. Appropriate measurements, such as using a patient global assessment, or taking a specific non-signal-joint measurement, should be included to capture treatment effects at other OA sites. This draft guidance uses knee or hip OA as its exemplar. Sponsors interested in approaches different in principle should consult the Agency before proceeding.

OA assessment to date is largely empirical and, to a great extent, patient driven. When deciding whether an OA product works, one asks certain obvious questions, such as: How much pain do you have? What can you do? Pain and function seem obvious, important, and interrelated assessments. It has even been argued that they should not be separated, even in principle. There is extensive experience and considerable understanding of the measurement of pain and function in OA, and of how the measurements depend on which instrument is used. It is equally obvious that loss of structural integrity is in some generic way the physical correlate of these symptoms. Uncertainty as to whether measurements of JSN alone will adequately reflect what *structure* connotes, or whether other metrics of structure should also be considered, is the fundamental concept underlying the formulation of the structure claim below.

A. Pain and Function

Traditionally, pain has been measured using the Likert or 10cm VAS scale,³ and both have been extensively validated in OA. Function has been more difficult to measure.

³ There is no important reason to prefer one scale over the other.

Draft — Not for Implementation

Because pain and function are interdependent and may not even be separable in principle, their simultaneous assessment is probably best done with instruments that measure both. With this in mind, two self-administered questionnaires have been developed, the Lequesne index, and the Western Ontario and McMaster Universities (WOMAC) osteoarthritis index. They assess pain, function, and stiffness in the knee and hip of OA patients. Both knee and hip are encompassed by the WOMAC, whereas there is one Lequesne questionnaire for the knee and a separate one for the hip. These questionnaires are useful because they contain more information content, compared to VAS or Likert. The derivation of the WOMAC relied heavily on patient feedback, whereas that of the Lequesne was more driven by physician judgment, but both questionnaires have been extensively validated in OA, including in surgical settings (knee and hip arthroplasty). Their metric characteristics do not differ in any important way. When using the WOMAC for sample size calculations, only disaggregated pain and function data are currently available.

Patient global assessments (also using a 10cm VAS or Likert scale) are overall measurements of whatever the patient deems most important. These are simple, unidimensional measurements, but they permit individualization of content and have been extensively validated in OA. The question remains as to whether patient global measurement should have a co-primary end-point status, similar to pain and function measurements.

B. Structure

Structure is a critical component of OA assessment, but the relationships between structure and pain and/or function and between structure and future outcomes (e.g., arthroplasty) are not well developed. At a minimum, cartilage destruction is a necessary but not sufficient prerequisite for arthroplasty. Additionally, OA may be asymptomatic early on, complicating the relationship between structure and outcome. Numerous epidemiological studies are underway to try to clarify these relationships.

Because OA is a disease of all the tissues in the joints, not just the cartilage, measurements of *structure* need to be seen broadly and capture important anatomic features, such as osteophytes or ligamentous instability, in addition to cartilage loss. There is evidence that osteophytes may be a major determinant of pain. However, the only well-characterized structure measurement currently available is radiographic JSN of the knee or hip. Because it is widely (but not universally) accepted that JSN improvement implies cartilage preservation and will thus be reflected in clinical benefit, measurement of the JSN using the x-ray is an appropriate structure measurement, subject to certain conditions (see below). In the future, measurement of cartilage volume using MRI may be able to replace the x-ray measurement of JSN, but currently technical problems remain, and adequate studies correlating MRI with x-rays are not yet available.

Determining what change in JSN of the knee or hip is *clinically relevant* to the patient

Draft — Not for Implementation

with OA is fundamental, but currently unknown. It may be that the clinically relevant change depends on disease duration or on baseline JSN. Ongoing epidemiology or future randomized trials may help describe what is clinically relevant. Nonetheless, this draft guidance proposes to use the concept *clinically relevant JSN* in determining what parallel clinical evidence should be used to demonstrate this claim (see below).⁴

V. OSTEOARTHRITIS CLAIMS

Labels inform prescribers and patients about the documented risks and benefits of a product. The two claims discussed in this section are *symptomatic treatment of pain and function* and *delay in structural progression*. The concept *OA prevention* is also introduced. Efficacy, for evidence in support of a claim, is demonstrated based on *statistically significant improvement compared with control*, and on a satisfactory overall risk-benefit analysis. A *delay in structural progression* claim may be contingent in some instances on parallel clinical evidence (see below).

Submissions may consist of combination products, separately targeted at pain or function, but the overall claim for *symptomatic treatment* cannot be subdivided. Other claims, such as *delay in time to joint surgery* are also possible in principle. The labeling should include descriptions of the trials, patients, durations, and results for all claims granted.

A. Treatment of Symptoms: Pain and Function

Efficacy endpoints in these trials should include measurement of pain, a patient global assessment, and a self-administered questionnaire (WOMAC or Lequesne), plus a measurement of structure if the trial lasts a year or more. X-rays are essential for risk-benefit assessment, even if no structure claim is sought or envisioned. The effects on non-signal joints (e.g., contralateral knee/hip, or hand OA) and the effects of confounders (osteophytes, rescue medication, assistive devices) should be standardized in the protocol and in the analysis. The effects on pain and function should be disaggregated, and each separately analyzed and noted in the proposed labeling. Products that disproportionately affect pain compared to function (or vice versa) will be considered approvable if the pain relief is large enough to yield overall success.

Trial duration for demonstration of symptom improvement should be at least three months; product- or device-specific considerations (e.g., new classes of agents, agents with delayed onset) may lengthen the duration. For example, a device may alter JSN immediately, and, in this case, evidence should include subsequent effects on adjacent tissues and on symptoms. If enough experience already exists with other products in the same class (as with e.g., nonsteroidals) to complement risk-benefit assessments, trials

⁴ The EMEA guidance deals with structure in a similar manner, January 29, 1998, p. 2, (<http://www.ema.europa.eu/udra.org/emea.html>).

could be as short as six weeks.

B. Delay in Structural Progression

The structural measurement currently proposed is demonstrating a slowing in the loss of knee or hip JSN using x-ray; other validated structural measurements may be developed in the future. Whether parallel symptom evidence should be included in the claim depends on what JSN outcome is achieved (see below), but symptom endpoints (using measurement of pain, a patient global assessment, a self-administered questionnaire) should be collected regardless of the outcome anticipated. Trials to demonstrate structure improvement should last at least one year. The reason for this is that the concept *structural improvement* connotes an element of durability, even if future technology allows the demonstration of slowing of the loss of JSN in shorter time periods. At present, the imprecision of the JSN measurement often results in trials lasting even longer than one year.

At present, few data speak to the validity or likelihood of a product showing benefits in delaying structural progression, but not showing benefits in improving patient symptoms. Although most products affecting inflammation would not be expected to slow JSN without affecting symptoms, it is possible that certain classes of products developed in the future may do so. A claim of slowing JSN (i.e., showing structural improvement) might plausibly be dissociated from other claims when the mechanism of action of the product, and/or the size of the effect on slowing of JSN, are suggestive of future clinical benefits. In general, products will not be considered for approval or for separate claims if (1) they are not anticipated to have different effects on these parameters, or (2) they show only small improvements in JSN without demonstrated effects in symptoms. Trials of agents expected to show isolated benefits should be carefully designed to preserve type 1 error. In addition, measurements of symptoms should be collected in all trials regardless of expectations of effects on JSN, because their assessment is critical for the analysis of the overall risks and benefits of the product.

A hierarchy of claims for structural outcomes is shown here.

1. *Normalize the x-ray.* An x-ray that shows a normalization of JSN is possible, at least in principle, and it would be the most convincing outcome of an improvement in structural integrity. But given our current understanding of OA, this outcome does not seem attainable for any currently studied class of products.
2. *Improve the x-ray.* An x-ray that shows a reversal in the JSN (i.e., a widening of the joint space) at endpoint compared to baseline would reflect new or regrown cartilage (and not the cartilage hypertrophy sometimes seen early in OA). This outcome would be convincing and require no formal parallel evidence of improvement in clinical outcomes.
3. *Slow JSN by at least a prespecified amount.* The amount of slowing of JSN to

Draft — Not for Implementation

demonstrate improvement of patient symptoms or function (i.e., the amount *clinically relevant*) remains unknown. Given that there exist important questions in this area, sponsors wishing to claim that their product slows JSN, but does not reduce symptoms should contact the Agency to discuss such a proposal, including the biological rationale, the relative amount of slowing of JSN they anticipate, and plans for studying long-term clinical outcomes. In general, sponsors seeking this claim should anticipate relatively large changes (<50 percent) in slowing JSN relative to the control arm.

C. Prevention of OA

Because the claim *prevention of the occurrence of OA*, using symptomatic and radiographic criteria, in new joints in patients with OA or in individuals at risk to develop OA in the future could be possible in principle, it is mentioned here. However, the practicality of this outcome would be challenging because OA can at times present radiographically first, and at times clinically first. Any trial used to demonstrate this outcome should first define the term *new OA*. Because one cannot repeatedly survey radiographically all possible OA sites, there are important unresolved assessment issues for designs capable of properly validating this claim. Furthermore, because this claim would be a chronic disease claim different in kind from all past approvals, it should have a more extensively and more formally documented safety database than previous submissions.

VI. TRIAL DESIGNS AND ANALYSES

Valid OA measurements are critical to good trial design, and for results to be convincing, the protocol should capture the important clinical extent of the disease. There is broad, international consensus that short-term OA assessment means measuring pain, performing a patient global assessment, and measuring function, and for long-term, a measurement of structure should also be included. If the patient global assessment is not construed as encompassing non-signal joints, a separate assessment should be performed to address effects on non-signal joints. Accordingly, the Agency suggests that all OA trials include a measurement of pain, a patient global assessment, use of a self-administered questionnaire, and if the trial lasts a year or more, a measurement of structure. If short-term controlled trials are extended unblinded, an x-ray assessment after one year may provide very useful information.

Specifically, the Agency is interested in public input as to whether more guidance should be provided about the use of the patient global assessment, or other approaches, to capture effects to the non-signal joints.

Standardization of the assessment of important covariates (assistive devices, rescue analgesics, osteophytes) is important, and an explanation and justification of how missing data was handled is critical. Excessive missing data may prevent the drawing of valid conclusions from the results. For either symptom or structure improvement claims, analyses can use statistical tests that compare mean changes from baseline if statistical normality can be assumed, or use methods for

Draft — Not for Implementation

ordered categorical data. The analytic plan should require prospective agreement on the adjustment for multiplicity if more than one endpoint is used (e.g., if a trial is pursuing both a symptomatic and structural endpoint). The protocol should specify whether secondary analyses will be used inferentially (i.e., whether their results will contribute to the submitted evidence base). This will always be the case where the trial is assessing both symptoms and structure, even if the development plan is not pursuing both a symptom and a structure claim. The analysis plan should adjust the p value used for the primary and the secondary analyses to maintain the same overall protection against false positive-results, and there should be preagreement on what analyses are hypothesis generating, needing no adjustment.

An alternative to the analysis of mean changes is the use of a categorical rating (akin to the ACR20 in RA) or a protocol-defined by-patient composite (e.g., *better, unchanged, or worse*). If well thought through, and especially if based on previous data, by-patient composites are arguably superior to analyses of means. However, composites may be difficult to prospectively define, and they sometimes engender loss of information.

In general, certain trial designs and certain endpoints mandate certain analyses and may preclude others. If a trial is to be persuasive in the end, its analysis should be prespecified, clear, and uncontroversial. As a result, the endpoint definitions and statistical tests that will be used should be considered carefully at the design stage. Endpoints should be evaluated both by how compelling they are to the clinician and what information loss occurs with their use. Similarly, statistical tests should have their underlying assumptions and their face-value persuasiveness assessed. Some aspects of relating the endpoints to the statistical tests are straightforward. For example, the statistical method should take into account whether the endpoint is continuous, binary, or categorical. However, the process of selecting endpoints and statistical tests is to some degree discretionary, so it is important to realize the effect that the content and interpretability of endpoints and statistical tests will have on the overall persuasiveness of the results. Finally, as a matter of good science, endpoints and analyses should be fully and prospectively specified (and always before unblinding).

VII. ASSEMBLING THE EVIDENCE

Symptom and structure claims can be pursued in the same trial or demonstrated in separate trials. Because the persuasiveness of trials showing a difference is generally greater than that of equivalence trials, it is highly desirable for a claim to be demonstrated in at least one trial showing superiority of the test product to placebo, to a lower dose of the agent, or to an active control.

VIII. OVERALL RISK-BENEFIT ASSESSMENT

Approval is predicated on controlled evidence demonstrating efficacy and an acceptable overall risk-benefit assessment overall. The size of the safety database at approval should be consistent

Draft — Not for Implementation

with the recommendations made by the International Conference on Harmonisation, but particular attention should be paid to systematically collecting information on known or suspected pharmacologic effects that might lead to delayed toxicities. Finally, if there is concern about rare but serious adverse events (e.g., from the mechanism of action or experience with similar agents) and if the safety database is small, a phase-4 safety monitoring program may be called for.