Chapter 6
Trial Design and Outcomes in Osteoarthritis

Nigel Arden

Introduction

As we continue to learn more about the pathophysiology of osteoarthritis, the number of potential new treatments will continue to increase, as will the need for research into its management. Over the past couple of decades, many excellent clinical trials in osteoarthritis have been performed; however, the study design and presentation of many trials has been inconsistent and often suboptimal. This leads to considerable difficulties when trying to assess the benefits of an individual treatment by assimilating a number of studies or when trying to make a comparison across different treatments. Some of these limitations reflect the design of the study (too small, patient selection, outcome measures), and others simply represent the presentation of the results in publications (details of study design, effect sizes).

In terms of the practicing clinician, the greatest limitation of clinical research into osteoarthritis is that the majority of studies are designed to assess whether a treatment works in a relatively homogenous group of patients with osteoarthritis. Clinicians want to know whether the treatment will work in a specific patient and therefore need to know the predictors of response to answer the more clinically important question of "in who does it work?" In this chapter, I will discuss the important steps that have been made by international cooperations to address these important issues.

Study Design

The traditional design for assessing the efficacy of a new treatment is a randomized, double-blind, placebo-controlled clinical trial. This design minimizes the chance of bias and will assess whether the new treatment is better than placebo. Once this has been established, a randomized, double-blind comparator trial should be performed to compare the efficacy of the new treatment with other existing therapies.

N. Arden
Senior Lecturer in Rheumatology, MRC Unit, Southampton General Hospital, Southampton, UK

D.M. Reid, C.G. Miller (eds.), *Clinical Trials in Rheumatoid Arthritis and Osteoarthritis*,

There are several forms of randomized controlled trials (RCTs). The favored design is a parallel group trial whereby participants are allocated to a treatment group for the whole study period. An alternative is the crossover design whereby each participant receives both treatments in a random order. This has the advantage of requiring a smaller number of patients to detect an effect. There are, however, several problems with this design, the most important is that if drugs have a prolonged or permanent effect, the subjects have to have a "washout period" between treatment arms to avoid a carryover effect; this design is therefore not suitable for trials of slow-acting symptom-modifying drugs or structure-modifying drugs. The N of 1 trial where patients receive the study drugs in a random order on more than one occasion has been proposed but suffers from the same limitations as the crossover design.

Osteoarthritis is rarely treated by a single treatment modality, and therefore there is increasing interest in the factorial parallel group RCT. This design is often used to assess a combination of two different treatments (A + B); there are four groups receiving treatments as shown in Fig. 6.1. This will answer several questions: Are treatments A and/or B better than placebo? Is either treatment superior to the other? Are both treatments together better than either alone?

There are several treatments for osteoarthritis where it is either impossible or unethical to blind patients or to use placebos; these include treatments such as surgery, education, or some forms of exercise therapy. In these situations, other designs can be used including randomized but not blinded parallel-group trial studies or occasionally in the case of total joint replacement observational studies such as cohort studies may need to be used.

Scientific versus Pragmatic Trials

An important early decision when designing a clinical trial is whether it should be a scientific or a pragmatic study. A scientific study design is used early on in the development of a treatment to assess its efficacy, whereas pragmatic designs are used later in development to asses the effectiveness, cost-effectiveness, and clinical

The factorial Design	
Group 1	Pl A + Pl B
Group 2	A + Pl B
Group 3	Pl A + B
Group 4	A + B

A = Drug A
Pl A = placebo to drug A
B = Drug B
Pl B = placebo to drug B

Fig. 6.1 The factorial design. A, drug A; Pl A, placebo to drug A; B, drug B; Pl B, placebo to drug B

Table 6.1 Scientific and pragmatic trials

	Scientific	Pragmatic
Measures	Efficacy	Effectiveness
Scientific validity	Good	Limited
Generalizability	Limited	Good
Patient selection	Well defined and homogenous	Representative of clinical population
Control group	Often placebo	Often usual clinical care + placebo
Exclusion criteria	Many	Few
Identify predictors of response	Limited	Good
Concomitant analgesics	Usually restricted	Usually unlimited
Intra-articular injections	Not allowed	Often allowed
Cost-effectiveness analyses	Limited use	Useful

predictors of response. Table 6.1 highlights the important differences between the two designs.

Study Duration

The duration of follow-up will depend on the therapy being assessed and the primary outcome measures. For an analgesic or nonsteroidal anti-inflammatory (NSAID) trial, 6 weeks would be the minimum duration to demonstrate its efficacy and adherence. For slower-acting symptom-modifying drugs and intra-articular therapies, 3 to 6 months would be required to demonstrate efficacy although longer may be required if assessing cost-effectiveness. A duration of 3 years is optimum for studies of structure-modifying drugs; although shorter periods can be used, it will require a greater number of patients to achieve the same statistical power.

Patient Selection

Scientific trials, early in a treatment development stage, tend to recruit well-defined patients and exclude patients with comorbidities in an effort to reduce the size and therefore costs of the study. A further technique often used in NSAID trials is the flare design: to enter the study, participants have to be on a NSAID, which is discontinued at the screening visit. Only those whose pain flares by a predetermined level are entered into the study, therefore including only patients who are responsive to NSAIDs. Whereas this is undoubtedly a scientifically valid and cost-effective approach to trial design, it induces several limitations. The results are not generalizable to the whole population of patients with osteoarthritis and more importantly neither is any estimation of effect size, NNTs, or cost effectiveness.

For trials of structure-modifying agents, symptomatic patients are usually selected on their radiographic grade. Traditionally, the Kellgren and Lawrence

Table 6.2 Kellgren and Lawrence grading of osteoarthritis

(a) Radiologic features on which grades were based

1. Formation of osteophytes on the joint margins or, in the case of the knee joint, on the tibial spines.
2. Periarticular ossicles; these are found chiefly in relation to the distal and proximal interphalangeal joints.
3. Narrowing of joint cartilage associated with sclerosis of subchondral bone.
4. Small pseudocystic areas with sclerotic walls situated usually in the subchondral bone.
5. Altered shape of the bone ends, particularly in the head of the femur.

(b) Radiographic criteria for assessment of osteoarthritis

Grade 0	None	No features of osteoarthritis
Grade 1	Doubtful	Minute osteophyte, doubtful significance
Grade 2	Minimal	Definite osteophyte, unimpaired joint space
Grade 3	Moderate	Moderate diminution of joint space
Grade 4	Severe	Joint space greatly impaired with sclerosis of subchondral bone

grading scale (Table 6.2) [1] has been used for knee osteoarthritis studies with grades II and III commonly being included. As most trials used JSN as the main outcome measure, a minimum joint space width is usually added into the inclusion criteria.

There is increasing interest in performing clinical trials in patients with knee pain or knee osteoarthritis defined by clinical criteria (Table 6.3) [2] without performing knee radiographs. Up to 50% of patients diagnosed with knee osteoarthritis in clinical practice will not fulfil the above radiographic criteria and are usually excluded from clinical trials. This therefore limits the generalizability of current clinical trials to a large proportion of patients in practice. These inclusion criteria are not suitable for structure-modifying drugs or early phase II or III studies of symptom-modifying drugs but are particularly suited to interventions such as home exercise regimens or phase IV trials of symptom-modifying drugs.

Control Arm

The choice of control groups will vary according to the trial question. In an early phase II or III study, it will invariably be a placebo group. However, it is becoming increasingly difficult to justify a placebo arm ethically when proven treatments are available. Trials of symptom-modifying drugs are therefore often performed against a comparator drug, but if against a placebo, participants are allowed free access to analgesics such as paracetamol. The use of free access to analgesia may introduce a conservative bias, as increased usage in the placebo group will minimize any treatment effect. To minimize this bias, participants are often asked to exclude escape analgesia for 48 hours before each assessment; however, this is becoming increasingly difficult to justify to an ethics committee. The alternative is to measure

Table 6.3 American College of Rheumatology (ACR) criteria for osteoarthritis of the hand, hip, and knee

	Clinical	Osteoarthritis is present if the items present are:
Hand	1. Hand pain, aching, or stiffness for most days or prior month. 2. Hard tissue enlargement of ≥2 of 10 selected hand joints.* 3. MCP swelling in ≤2 joints. 4. Hard tissue enlargement of ≥ 2 DIP joints. 5. Deformity of ≥1 of 10 selected hand joints.	1, 2, 3, 4 *or* 1, 2, 3, 5
	Clinical and radiographic	
Hip	1. Hip pain for most days of the prior month. 2. ESR ≤ 20 mm/h (laboratory). 3. Radiograph femoral and/or acetabular osteophytes. 4. Radiograph hip joint-space narrowing.	1, 2, 3 *or* 1, 2, 4 *or* 1, 3, 4
	Clinical	
Knee	1. Knee pain for most days of prior month. 2. Crepitus on active joint motion. 3. Morning stiffness ≤30 minutes in duration. 4. Age ≥38 years. 5. Bony enlargement of the knee on examination.	1, 2, 3, 4 *or* 1, 2, 5 *or* 1, 4, 5
	Clinical and radiographic	
	1. Knee pain for most days of prior month. 2. Osteophytes at joint margins (radiograph). 3. Synovial fluid typical of osteoarthritis (laboratory). 4. Age ≥40 years. 5. Morning stiffness ≤30 minutes. 6. Crepitus on active joint motion.	1, 2 *or* 1, 3, 5, 6 *or* 1, 4, 5, 6

MCP, metacarpophalangeal; DIP, distal interphalangeal; ESR, erythrocyte sedimentation rate.
*Ten selected hand joints include bilateral second and third DIP joints, second and third proximal interphalangeal (PIP) joints, and first carpometacarpal (CMC) joint.

use with an analgesia diary with this data then used as a secondary outcome measure and also as a covariate in multivariate statistical analyses of the primary outcome measure.

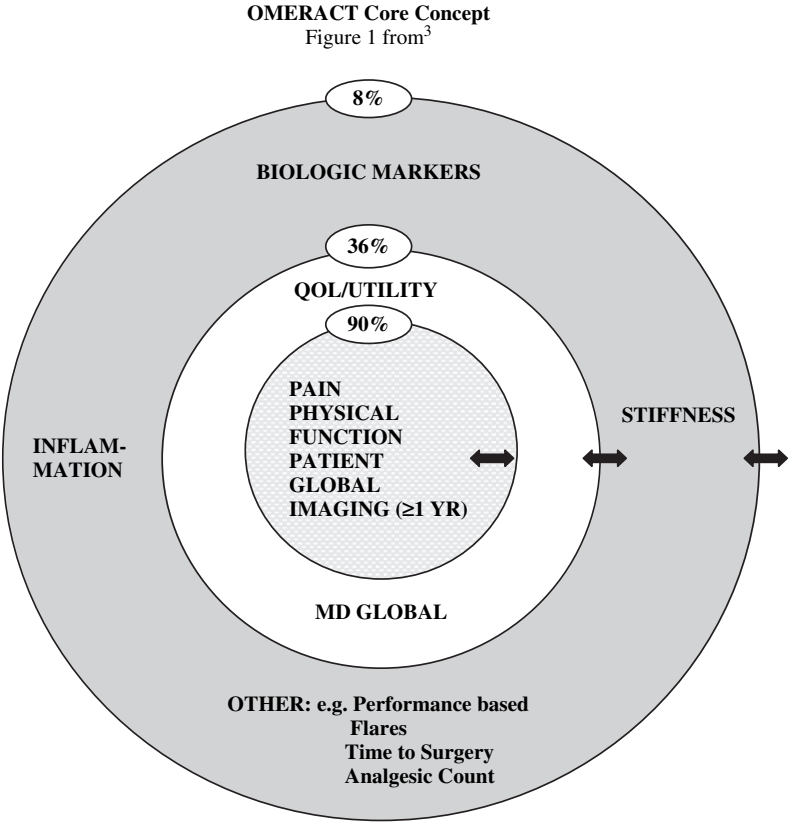
For structure-modifying drugs, where there is little evidence for structure modification with currently available agents, it is still acceptable to use a placebo. A bigger issue for these studies is whether participants taking glucosamine sulfate or chondroitin sulfate should be excluded from these studies, because of their proposed structure-modifying effects. As the prevalence of usage increases, this will become an ever-increasing problem because of problems of recruitment and also of the generalizability of any study that does not include them. One option is to include participants on a long-term stable dose but to stratify recruitment according to usage.

Sample Size

The determination of the sample size is a crucial step in the design of a clinical trial. It is essential that the study recruits enough patients to definitively determine the efficacy of the intervention, but ideally it should be large enough to also detect predictors of response to treatment. It is important to allow for dropouts from the study, which can be as high as 25% in studies of more than 12 months duration. To perform a sample size calculation, it is important to set the type I (α) and type II (β) error rates; the standard type I error is 0.05 and the type II error should ideally be 0.10, which means that the study has 90% power to detect the specified effect. There is still little agreement on how to define the difference between the studied treatments that the study should aim to detect. The Outcome Measures in Rheumatoid Arthritis Clinical Trials (OMERACT) group has considered this issue in some detail and come up with several options [3]. There are a number of options that use the minimum statistically detectable difference of the outcome tool used. More clinically useful are definitions based on clinical improvement, including the minimum perceptible clinical improvement (MPCI) and the minimum clinically important difference (MCID) [4]. OARSI has recently published responder criteria [5], which may in the future be used to perform dichotomous sample size calculations.

Outcome Assessment

Outcome measures used in clinical trials need to be valid, reliable, and responsive to change. OMERACT has defined a core set of outcome measures that should be measured in all osteoarthritis trials, with a list of additional optional measures [3] (Fig. 6.2). Measures of pain, physical function, and patient global assessment should be measured as outcome measures for all clinical trials. A list of commonly used instruments for the assessment of each of these measures is shown in Table 6.4. Imaging of the index joints should be performed in all studies of 12 months or greater; as an outcome measure for studies of structure-modifying drugs; and as a safety measure for other studies.



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

Fig. 6.2 OMERACT core concept (From Bellamy N, Kirwan J, Boers M, Brooks P, Strand V, Tugwell P, et al. Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip, and hand osteoarthritis. Consensus development at OMERACT III. J Rheumatol 1997;24:799–802.)

The use of additional analgesics should be recorded and used as a secondary outcome measure for all studies in which they are allowed. Studies assessing cost-effectiveness should also record usage of all other medications and therapists in order to obtain a total costing to the health service. In studies of structure-modifying

Table 6.4 Commonly used instruments for the assessment of measures

	Pain	Stiffness	Function	Health status
Knee and hip	WOMAC [7] Lequesne	WOMAC Lequesne	WOMAC Lequesne [8]	SF-36 EuroQol [9]
Hand	AUSCAN [10]	AUSCAN	AUSCAN Cochin [11]	SF-36 EuroQol

drugs, joint surgery should be recorded during the study, and preferably with long-term follow-up after the study has finished.

Radiographs of the index joint are currently the primary outcome measure of choice for studies of structure-modifying drugs. Change in joint space width is the preferred measure of disease progression for hips and knee osteoarthritis, and much work has been performed to define the most reproducible technique of assessing width (see Chapter 14). Although there is not yet agreement on a single technique, there is general consensus that for the knee, the patient should be weight bearing with the knee in a semiflexed position. Much work is being performed to assess the role of MRI as an outcome measure, however currently it cannot be recommended as the sole primary outcome measure. There is currently much interest in the role of biochemical markers of cartilage and synovial turnover as outcome measures, however although the results look interesting, they are not yet suitable as a primary outcome measure.

Compliance

It is essential to know if the participants are compliant with the study protocol. What used to be termed *compliance* is now split into two separate entities: *continuance* is whether the patient remains on the treatment regimen, and *adherence* is the degree to which the patient adheres to the regimen (i.e., how many of the prescribed tablets do they take). Adherence is usually regarded as acceptable if the participant takes at least 80% of the prescribed medication. There are several methods of measuring adherence: drug diary, tablet counting, direct observation of ingestion, and plasma monitoring. The last two techniques are not suitable for large-scale clinical trials, and most studies will use one of or both diaries or tablet counting.

Statistics

The intention to treat (ITT) analysis, where all patients analyzed are included in the final analysis irrespective of whether they were adherent or completed the trial, is the traditional method of analyzing clinical trials as it is the least susceptible to bias. There are several methods for dealing with noncompleters; there is, however, no consensus on which is the best. The most commonly used is the last observation

carried forward technique, where the last observation recorded is used as the study end point. Another techniques is to use the best, worst, and last observation carried forward to give a spread of possible end points within which the true value should lie. More recently utilized are imputation techniques, whereby the last observation measured is used to predict the final result based on other participants in the trial. A completers analysis can be performed in participants fully adherent to the protocol to give an estimate of the best effect of the intervention. Although this can provide useful information, it should always be accompanied by an ITT analysis.

Many studies measure outcomes at several time points providing the opportunity for more complex and informative statistical approaches. It is possible to use techniques such as repeated measure analysis of variance (ANOVA) to measure all time points to give an idea of the cumulative effect of the intervention over the whole study period. For dichotomous outcome, it also allows the use of survival analyses, which allow for the time to event rather than just the presence of the event.

Presentation and Dissemination

For a clinical trial to change clinical practice, it must be well presented to provide all of the information that the reader or health care provider requires. The Consort guidelines detail the basic requirements for publication and have been adopted by the majority of medical journals [6]. It is important that all clinical trials are published, including negative trials, to avoid publication bias.

Study Organization

It is important when designing the study protocol to include all of the professions that will be involved at the very beginning. As well as rheumatologists and orthopedic surgeons, this should include physiotherapists, general practitioners, statisticians, and health economists. It is very important to perform a pilot study before finalizing the protocol of the study. This should test all of the proposed outcome instruments used in the study and the logistics of the day-to-day operation. Most importantly, it should address the issue of recruitment rates, which is the most common downfall of large studies. With increasing legal requirements relating to the conduct of clinical trials, it is important to appoint an independent data monitoring and ethics committee in the early stages of the trial in addition to the trial steering committee.

Conclusion

The past two decades have seen a major advance in the design of clinical trials in osteoarthritis. This is largely due to the standardizing of outcome measures and trial reporting. This will allow for greater evaluation of existing and novel therapies for osteoarthritis in the future.

References

1. Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthrosis. Ann Rheum Dis 1957;16:494–502.
2. Cooper C. Osteoarthritis and related disorders: epidemiology. In Klippel JH, Dieppe PA, eds. Rheumatology. London: Harcourt Publishers Ltd; 2000:8.2.1–8.2.8.
3. Bellamy N, Kirwan J, Boers M, Brooks P, Strand V, Tugwell P, et al. Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip, and hand osteoarthritis. Consensus development at OMERACT III. J Rheumatol 1997;24:799–802.
4. Bellamy N, Carr A, Dougados M, Shea B, Wells G. Towards a definition of "difference" in osteoarthritis. J Rheumatol 2001;28:427–430.
5. Dougados M, Leclaire P, van der HD, 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 Cartilage 2000;8:395–403.
6. The CONSORT Statement. 2004.
7. 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;15:1833–1840.
8. 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 Suppl 1987;65:85–89.
9. Hurst NP, Jobanputra P, Hunter M, Lambert M, Lochhead A, Brown H. Validity of Euroqol—a generic health status instrument—in patients with rheumatoid arthritis. Economic and Health Outcomes Research Group. Br J Rheumatol 1994;33:655–662.
10. Bellamy N, Campbell J, Haraoui B, Gerecz-Simon E, Buchbinder R, Hobby K, et al. Clinimetric properties of the AUSCAN Osteoarthritis Hand Index: an evaluation of reliability, validity and responsiveness. Osteoarthritis Cartilage 2002;10:863–869.
11. Poiraudeau S, Chevalier X, Conrozier T, Flippo RM, Liote F, Noel E, et al. Reliability, validity, and sensitivity to change of the Cochin hand functional disability scale in hand osteoarthritis. Osteoarthritis Cartilage 2001;9:570–577.