

Osteoarthritis and Cartilage



Review

OARSI Clinical Trials Recommendations: Design and conduct of clinical trials of rehabilitation interventions for osteoarthritis



G.K. Fitzgerald [†]*, R.S. Hinman [‡], J. Zeni Jr. [§], M.A. Risberg ^{||}, L. Snyder-Mackler [§], K.L. Bennell [‡]

[†] Department of Physical Therapy, School of Health and Rehabilitation Sciences, University of Pittsburgh, Pittsburgh, PA, USA

[‡] Centre for Health, Exercise & Sports Medicine, Department of Physiotherapy, University of Melbourne, Australia

[§] Department of Physical Therapy, University of Delaware, Newark, DE, USA

^{||} Norwegian Research Center for Active Rehabilitation, Department of Orthopedic Surgery, Oslo University Hospital, Oslo, Norway

ARTICLE INFO

Article history:

Received 3 December 2014

Received in revised form

6 March 2015

Accepted 9 March 2015

Keywords:

Rehabilitation

Clinical trials

Osteoarthritis

SUMMARY

A Task Force of the Osteoarthritis Research Society International (OARSI) has previously published a set of guidelines for the conduct of clinical trials in osteoarthritis (OA) of the hip and knee. Limited material available on clinical trials of rehabilitation in people with OA has prompted OARSI to establish a separate Task Force to elaborate guidelines encompassing special issues relating to rehabilitation of OA. The Task Force identified three main categories of rehabilitation clinical trials. The categories included non-operative rehabilitation trials, post-operative rehabilitation trials, and trials examining the effectiveness of devices (e.g., assistive devices, bracing, physical agents, electrical stimulation, etc.) that are used in rehabilitation of people with OA. In addition, the Task Force identified two main categories of outcomes in rehabilitation clinical trials, which include outcomes related to symptoms and function, and outcomes related to disease modification. The guidelines for rehabilitation clinical trials provided in this report encompass these main categories. The report provides guidelines for conducting and reporting on randomized clinical trials. The topics include considerations for entering patients into trials, issues related to conducting trials, considerations for selecting outcome measures, and recommendations for statistical analyses and reporting of results. The focus of the report is on rehabilitation trials for hip, knee and hand OA, however, we believe the content is broad enough that it could be applied to rehabilitation trials for other regions as well.

© 2015 Osteoarthritis Research Society International. Published by Elsevier Ltd. All rights reserved.

Introduction

A Task Force of the Osteoarthritis Research Society International (OARSI) has previously published a set of guidelines for the conduct of clinical trials in osteoarthritis (OA) of the hip and knee¹. Limited material available on clinical trials of rehabilitation in people with OA has prompted OARSI to establish a separate Task Force to elaborate guidelines encompassing special issues relating to rehabilitation of OA. The Task Force was composed of a multi-national group of six academic physical therapists with expertise in OA, all

of whom have extensive experience in designing and conducting clinical trials in rehabilitation of people with OA.

The Task Force identified three main categories of rehabilitation clinical trials. The categories included non-operative rehabilitation trials, post-operative rehabilitation trials, and trials examining the effectiveness of devices (e.g., assistive devices, bracing, physical agents, electrical stimulation, etc.) that are used in rehabilitation of people with OA. In addition, the Task Force identified two main

* Address correspondence and reprint requests to: G.K. Fitzgerald, Department of Physical Therapy, University of Pittsburgh, Bridgeside Point 1, 100 Technology Drive, Suite 210, Pittsburgh, PA 15219-3130, USA. Tel: 1-412-383-6643; Fax: 1-412-648-5970.

E-mail address: kfitzger@pitt.edu (G.K. Fitzgerald).

Summary Box: Design and conduct of clinical trials for rehabilitation in people with osteoarthritis

Categories of Clinical Trials

- Non-operative rehabilitation
- Post-operative rehabilitation
- Effectiveness of devices (e.g., assistive devices, bracing, physical agents, electrical stimulation, etc.)
- Outcomes related to symptoms and function
- Outcomes related to disease modification

Entering Patients into the Trial

- Provide a thorough description of the participants (e.g., demographics, how the diagnosis of OA was determined, comorbidities, baseline outcome measures)
- Describe methods of patient recruitment for the study
- Describe the inclusion and exclusion criteria

Conduct of the Study

- Overview of the study design
- Description and rationale for selection of control/comparator groups (e.g., placebo vs comparator intervention)
- Thorough description of interventions so that others can replicate the interventions (algorithms for treatment selection/progression, dosage (intensity, frequency, duration), adherence strategies, home programs, training of treatment providers and treatment fidelity methods)
- Blinding procedures
- Managing and recording adverse events

Outcome Measures

- Should include reliable and valid measures of pain, patient-reported function and disability, performance-based measures of function, patient global assessment
- Explanatory trials should include measures to confirm hypothesized mechanisms of treatment effect
- Results of outcome measures should be benchmarked with known age and/or condition-matched normative values or previously established outcome values if available.
- The primary outcome should be expressed in terms of mean, standard deviation, and 95% confidence intervals. Authors should also consider reporting mean differences between groups and the standard deviation and 95% confidence intervals of the differences.

categories of outcomes in rehabilitation clinical trials, which include outcomes related to symptoms and function, and outcomes related to disease modification. The recommendations for rehabilitation clinical trials provided in this report encompass these main categories.

The recommendations do not specifically address each possible rehabilitation intervention that can be studied in clinical trials. In developing the recommendations, the Task Force attempted to make the descriptions of required elements broad enough to encompass a variety of rehabilitation interventions or approaches.

The Task Force considered rehabilitation to include interventions that intended to reduce symptoms, improve functional capacity, and/or promote healing with the goal of reducing or eliminating disability. Rehabilitation interventions were assumed to include therapeutic exercise, neuromuscular control and functional retraining, physical modalities, electrical stimulation and acupuncture, assistive devices, bracing and orthotics, self management education, strategies for improving self-efficacy, pain coping strategies, etc. It should be noted that while some elements of studies using therapeutic exercise are covered in these guidelines, recommendations for studies addressing the use of exercise and weight loss will also be covered in another article related to this special issue.

Each member of the task force was assigned a section of the guideline for development (e.g., entering subjects into the trial, study design, interventions, adverse events and protocol violations, outcome measures, and statistical considerations). Members of the task force reviewed each section and made comments and suggestions for revision. In instances where there was disagreement on a specific item or issue, continued discussion and search of the literature was used to resolve issues and a majority opinion was used for final decision.

Finally, the recommendations presented here were focused on hip, knee and hand OA. These guidelines do not necessarily include specific elements that might be needed in clinical trials addressing rehabilitation approaches for OA of the spine, shoulder, and foot and ankle. However, we believe that many of the recommendations in the present guidelines would likely apply to trials examining the effectiveness of rehabilitation interventions for these regions.

Entering patients into the trial

This section deals with aspects of the protocol related to recruitment and enrollment of participants into the trial and baseline data that will need to be obtained in order to clearly describe the study participants.

Demographics

Demographic information provides a description of the general characteristics of participants in the trial. Demographic information can help to determine if randomization procedures were successful in equating groups on basic characteristics that may have an impact on the study outcome. This information can also be useful in determining whether any of these characteristics might be useful predictors of clinical outcome associated with the study interventions.

Demographic characteristics should include participant age, sex, race/ethnicity, height, weight, body mass index (BMI), educational status, marital status, living arrangements, and employment status (full-time, part-time, unemployed).

Diagnosis of OA

The study protocol should explain how the diagnosis of OA was confirmed for study participants. This should be done using validated criteria for diagnosis of OA such as the American College of Rheumatology's (ACR) diagnostic criteria for hip, knee, or hand OA^{2–4}. For non-operative rehabilitation focusing on symptoms/function, established clinical criteria (without use of radiographs or laboratory data, such as the ACR clinical criteria) may be adequate particularly if the interventions may be beneficial regardless of whether there is radiographic evidence of OA. It should also be stated whether participants have unilateral vs bilateral involvement. For studies concerning post-operative rehabilitation or a pre-

operative rehabilitation trial aiming to influence post-operative results, the specific type of surgery or surgeries that will be targeted in the study should be clearly identified.

Radiographic/imaging severity

It may not be necessary for all rehabilitation studies to report disease severity through radiographic or other forms of imaging, particularly if the main outcomes of the study are only concerned with symptom or function modification. However, a description of disease severity through imaging procedures will be necessary if: (1) there is intention to include/exclude participants with radiographic OA involvement of a specific joint or joint compartment (e.g., medial tibiofemoral compartment, patellofemoral joint, *etc.*); (2) the effects of the rehabilitation intervention on structural disease is to be assessed; and/or (3) the potential of structural disease severity and/or extent of malalignment on mediation or modification of the intervention on the clinical outcome is to be evaluated. For example, trials investigating the efficacy of mechanical interventions that are targeted towards a particular joint compartment (e.g., unloading knee brace that aims to unload either the medial or lateral tibiofemoral joint) will usually require imaging for these reasons. In post-operative rehabilitation trials, imaging procedures may be required if participant inclusion/exclusion criteria are based on proper placement of surgical components, quantification of disease severity in the non-operative limb, or if an aim of the study is to examine the effects of rehabilitation on surgical components. If a description of imaging-based disease severity is to be included, then it should be reported using a validated severity quantification scale, following the recommendations reported in the article(s) in this special issue on imaging.

Co-morbidities/other medical history

Co-morbid conditions and concurrent interventions may influence the outcome of rehabilitation interventions and should be accounted for in the study protocol. Where concurrent interventions are concerned, providing as much detail about the application of the intervention would be important in determining its role as a potential confounder. In particular, for trials of devices (e.g., braces, orthoses, footwear, gait aids), current use of these interventions may be considered an exclusion criterion for trial enrollment due to difficulties administering the intervention of choice when the participant is already receiving a similar intervention. See [Table 1](#) for a list of information that should be obtained on study participants.

Table 1

Documentation of comorbidities and potential confounders of treatment outcome

-
- Chronic Diseases or conditions (e.g., diabetes, osteoporosis, heart disease, hypertension, neurological disorders)
 - History of previous injury or surgery related to the study target joint with OA
 - Other current pain sites
 - Past and current tobacco and alcohol use
 - Past and current medications for treatment of joint pain/OA
 - Past and current non-pharmacological treatments (e.g., exercise, orthotics, *etc.*) of joint pain/OA (**Providing as much detail as possible concerning the application of these interventions such as dosage and duration, setting in which the treatments were applied, whether they were active or passive interventions, could be helpful in determining their role as a potential confounder**)
 - History of most recent cortico-steroid and/or hyaluronan injection for target joint pain/OA and other joints
 - Previous use of study interventions
 - Subject expectations of treatment effectiveness
 - Occupational activity (sedentary, moderately physical labor, hard physical labor)
 - Recreational activity (sedentary, moderately physically active, highly physically active)
 - For post-operative rehabilitation trials:
 - Surgical complications or other intra-operative findings that may influence outcome measures
 - Was pre-operative rehabilitation provided
 - If study focus is on outpatient rehabilitation, report any inpatient or home-based therapy that may have been received
-

For post-operative rehabilitation trials, information concerning surgical complications or other intra-operative findings that may have an influence on the outcome of rehabilitation should be reported. If participants in a post-operative rehabilitation trial receive pre-operative rehabilitation, this should also be reported. When the focus of the post-operative rehabilitation trial is on outpatient rehabilitation, inpatient or home-based therapy the participant may have received before entering the outpatient trial should be reported.

Baseline symptoms, function, physical activity levels, health-related quality of life

Participants in rehabilitation studies should be described with respect to their baseline pain, function, physical activity, and health-related quality of life. Reliable and validated measures should be used for these purposes. Examples of reliable and valid measures for these constructs are provided in the outcome measures section below. If study interventions are designed to address specific impairments (e.g., joint range of motion, muscle strength, balance, aerobic capacity, joint proprioception, *etc.*) or if it is believed that the impairments may be treatment outcome modifiers or mediators, then those impairments should be measured at baseline using reliable and valid methods. If study interventions are designed to address specific impairments (e.g., joint range of motion, muscle strength, balance, aerobic capacity, joint proprioception, *etc.*) or psychosocial variables (e.g., depression, anxiety, fear of physical activity, pain catastrophizing, self-efficacy, *etc.*) or if it is believed that these may be treatment outcome modifiers or mediators, then these should be measured at baseline using reliable and valid methods. Other variables that might be potential confounders, or treatment moderators or mediators such as subject expectations of treatment effectiveness, occupational activity (sedentary, moderately physical labor, hard physical labor), and recreational activity (sedentary, moderately physically active, highly physically active) may also be useful to record at baseline.

Methods of recruitment

The methods used to recruit participants and the sources of recruitment should be clearly explained (e.g., use of registry lists, public advertisements/announcements, direct referral from healthcare providers, *etc.*)

Inclusion/exclusion criteria

Inclusion and exclusion criteria should be clearly defined. Participants may be included based on age limits, sex, diagnostic

criteria of OA, joint(s) or joint compartment being targeted. Studies examining post-operative rehabilitation need to be clear on the target surgical procedure(s) that will serve as the basis for inclusion into the study. Inclusion criteria for levels of baseline symptoms, function, and/or disability should be established to ensure that minimum clinically important differences can be detected at follow-up testing.

Exclusion criteria should also be clearly defined. Participants may be excluded based on age limits, sex, and presence of inflammatory arthritides (gout, RA, psoriatic arthritis, etc.). Comorbidities that serve as a basis for exclusion should also be described. It is important to note that, in some trials, just the presence of pain or pathology at other sites may be considered an exclusion criterion. For example, in a trial evaluating shoe orthoses for people with knee OA, concurrent pain and/or pathology in the foot or ankle may be a contra-indication to treatment with shoe orthoses and thus may be exclusion criteria. Previous or current treatments may also serve as a basis for exclusion and should be described. This may include surgical/medical procedures, pharmacological interventions, or specific previous or concurrent treatments (e.g., if testing a brace, exclude people who are currently using a brace or if testing the effect of strength training, exclude people who have been currently participating in a regularly scheduled strength training program). Participants may also be excluded on the basis of contraindications to the specific study interventions (e.g., if poor skin condition, then might exclude from a taping or bracing trial).

Conduct of the study

Study design

Studies should be randomised, controlled, and parallel in design. At a minimum, studies should be assessor-blinded, and ideally, participant-blinded as well, although it is acknowledged that this is often not possible in rehabilitation studies. Most clinical trials aim to test the superiority of one treatment over another, and thus are often termed superiority trials. However, although the randomized, controlled parallel design would be the gold standard in most interventions studies, different trials and designs could be considered depending on the aims of the study. For detailed descriptions of other trial design options, the reader is referred to the article in this special issue on research designs and statistical considerations.

Clinical trial designs may be categorized as “explanatory” or “pragmatic” trials⁵. Explanatory trial (or efficacy trial) designs should be used when investigators wish to determine whether an intervention produces the hypothesised benefit that it was intended to produce and therefore tests causal or mechanistic hypotheses⁵. Explanatory trials examine interventions under ideal (research) circumstances, with a higher emphasis on the internal validity of the research design. In contrast, pragmatic trial (or effectiveness trial) designs should be employed if investigators wish to measure whether a treatment is beneficial when tested and delivered in “real world” clinical settings⁵. Pragmatic trials are more lenient on control of design elements, such as clinician experience, deviations from intervention protocol, participant characteristics/comorbidities, etc. This is because these elements are not always well controlled in real clinical situations and the intent of the pragmatic trial is to determine how well the interventions may work under these less than ideal, yet realistic conditions. Pragmatic trial designs tend to emphasize external validity at the expense of some internal validity. It is important to recognise that explanatory and pragmatic trials exist on a continuum and in many cases a trial design may incorporate elements of both. Thus investigators need to consider what their intended aims

are for their study in order to determine where on this continuum they would like to have the trial design fall. For the remainder of this manuscript we will try to indicate where applicable how investigators may alter some design characteristics based on whether more explanatory or pragmatic designs are pursued.

Frequently for rehabilitation interventions, comparative effectiveness study designs (where two or more groups are randomly assigned to receive different interventions without knowing the research hypothesis) may be appropriate. Comparative effectiveness trials may be particularly useful for testing the effectiveness of novel treatments relative to established interventions, or when participant blinding is not possible. Occasionally, crossover trials may be appropriate. Crossover trials may be suitable for interventions where treatment effects are likely to occur relatively quickly and subside rapidly once the intervention is ceased and with little likelihood of long-term carry-over on the study outcomes (e.g., knee bracing, shoe orthotics, patellar taping, gait aids). Crossover trials should randomly assign the treatment sequence to participants and carry-over effects should be minimized by a sufficiently long “wash-out” period between treatments. Investigators may wish to determine whether an intervention is therapeutically similar or equivalent to another existing or accepted treatment with established efficacy. In these instances, equivalence or non-inferiority trials (as opposed to superiority trials) are the most appropriate trial design and it is critical that an a priori margin of equivalence or non-inferiority is determined. These types of study designs are only appropriate if the “new” treatment offers clear advantages or benefits (often in terms of cost or practical application) over the existing or reference treatment.

The study must be approved by the local human research ethics committee prior to recruitment commencing, and all study participants must provide informed consent to participate. The trial must be registered prospectively (prior to participant recruitment commencing) with an appropriate clinical trials registry (e.g., <http://clinicaltrials.gov/>, <http://www.anzctr.org.au/>, or <http://www.isrctn.com/>). In addition, investigators may wish to publish the trial protocol (e.g., *BMC Musculoskeletal Disorders*, *Journal of Physiotherapy*, *Trials*, *BMJ Open*, etc) to alert the scientific community that the trial is underway and to publish the protocol in sufficient detail that is often not possible when writing up the final results of the trial.

The study design should include a screening phase to ensure participants fulfill selection criteria and are eligible to enter the study. The screening phase may encompass one or more of telephone interview, self-reported questionnaires, radiographic review and/or a clinical examination. Employing multiple screening processes increases the likelihood that participants fulfill selection criteria and provides some reassurance that participants are likely to adhere to the trial protocol and its requirements, which may be particularly important in studies of rehabilitation. For trials of devices, such as braces or orthoses, where participant application of and tolerance of the device may influence adherence and trial drop-out rates, a “run-in” approach may be employed in the screening phase, where participants are provided with a similar device to the one being studied (i.e., an alternate type of orthotic so as not to unblind participants) and only those who are sufficiently adherent to use of the device over an initial short time-frame are enrolled. While such an approach increases the likelihood of maximizing intervention adherence and minimizing drop-outs in the trial, it does result in a participant sample that is less generalizable to the wider population and researchers should consider these drawbacks when designing their study. When pain is to be measured as a primary study outcome, and/or a minimum pain score is used as a study inclusion criteria, it may be helpful to screen baseline pain at several occasions. There can be clinically significant weekly

variation in pain reported by patients with OA^{6,7} and establishing a baseline in this manner may maximize the accuracy and reliability of baseline pain data.

Depending upon the number of outcomes being assessed, and the burden imposed by the assessment tasks on the participant (with respect to time constraints and physical demands), one or more assessment sessions may be required to collect all of the required baseline data. When participant-reported outcome measures (e.g., questionnaires) are being used, it may be appropriate for participants to complete the questionnaires at home and post them back, or online and submit them to investigators electronically. In these cases, investigators should check each questionnaire carefully for any incomplete items and ensure participants answer any incomplete questions. Wherever possible, it is desirable that the same assessor examines the same participant at each visit, including at follow-up assessments, to minimize unwanted variability in data. Ideally, and particularly when pain is assessed as a study outcome, participants should be re-assessed on the same day of the week and preferably at the same time of day over the course of the trial.

Control conditions (comparator groups)

Studies should always include a comparison group. The comparison group may be a no treatment control (i.e., participants continue with their usual care of their symptomatic OA which may involve drugs or other interventions but are not allocated to receive the active intervention being evaluated), an active comparator where the comparator has been previously shown to be effective, or a placebo (sham) treatment. It is also possible in rehabilitation trials to have the comparison groups include surgical intervention (and often times the post-surgical rehabilitation that accompanies surgical treatment) where the non-operative group might be the experimental group and the surgical intervention group serves as the active comparator.

Although a placebo-control is generally considered the ideal comparator group in drug trials for participant blinding in order to provide a less biased estimate of treatment effect, placebo interventions in rehabilitation trials are difficult to devise and implement. In fact, for complex interventions involving a package of care (e.g., physiotherapy, progressive individualised exercise programs that may also be combined with use of splinting, braces or orthoses) where characteristic (specific treatment effects) and incidental (placebo, non-specific, context factors) treatment effects are intertwined and often indistinguishable, the use of a placebo (sham) control may fail to detect the whole characteristic treatment effect and may lead to false negative results⁸. In these instances, pragmatic designs that utilise usual care or another active comparator may be more appropriate. When studies examine interventions that involve only the provision of a device (e.g., shoe orthotics, braces, splints, and/or footwear with minimal ongoing input from a health professional), the characteristic treatment effects are usually distinct from incidental effects and a placebo-control is often possible. An ideal placebo (sham) comparator should be indistinguishable from the active treatment, credible to the participants and inert with no specific therapeutic effects. For more information on the strengths and weaknesses of various comparator groups in clinical trials, the reader is referred to the article in this special issue concerning statistical and experimental design considerations.

Blinding

Wherever possible, the assessor, participant and therapist should be blinded to which group is receiving the experimental treatment in order to minimize the potential risk of bias (e.g.,

performance and response bias). However, it is often difficult to achieve participant blinding in rehabilitation trials, especially when participants of both treatment arms are being treated in the same setting and/or by the same health care providers. Furthermore, under most circumstances it is not possible to blind the therapist who is applying the treatment. Given that meta-analyses show that inadequate allocation concealment and lack of blinding are associated with over-optimistic estimates of treatment effects for patient-reported outcomes⁹, efforts to minimize bias in trials with participant-reported outcomes are particularly crucial.

The assessor should always be blinded in a clinical trial and this is almost always possible in rehabilitation trials. In rare cases where assessors cannot be blinded, it is important that objective and reliable outcome measures are utilized to minimize risk of bias. Duplicate assessment of outcomes by two independent assessors and reporting of the level of agreement achieved by the assessors should also be considered in these instances. In trials where patient-reported outcome measures are used, and the patient is blinded to treatment allocation, the assessor is considered to also be blinded. Irrespective of participant and assessor blinding, statistical analyses should ideally be performed by a blinded statistician/investigator.

Although ideal, blinding of the treating therapist is almost always impossible in rehabilitation studies, except for in studies of electrotherapy where the device delivering the intervention may be custom-developed to deliver the intervention (or not) based on the input of a randomization code by the blinded therapist. [Table II](#) provides suggested strategies that may be employed to minimize bias in trials where therapist and/or participant blinding is not possible.

When reporting the results of the trial, it is important to clearly state who was blinded (e.g., participants, assessors, therapists). In addition, the methods used for masking must be clearly stated. It is possible that, in addition to a blinded assessor (to assess the participant on primary outcomes regarding efficacy and adverse events), an unblinded assessor may be needed (e.g., to measure biomechanical mediators of effect whilst participant wears their allocated mechanical intervention that is visually identifiable as the active intervention (e.g., orthotics, knee brace, footwear)) in some studies.

Interventions

Considerations around the interventions and their delivery will be influenced by the position of the trial design along the continuum from explanatory to pragmatic. In rehabilitation trials, the experimental intervention can range from a single modality through to a complex package of care. While it may be ideal to provide standardized treatment in a clinical trial, this is not necessarily reflective of clinical practice in which rehabilitation is individualized taking into account patient presentation on clinical examination, patient preferences and past history of treatment. Furthermore, it does not align with clinical treatment guidelines for OA where individualized treatment is recommended due to different subgroups of OA patients as well as varying clinical presentations¹⁰. A trial that allows greater flexibility in applying the treatment will lie closer to the pragmatic end of the trial design spectrum whereas a standardized treatment with strict instructions for every element will be closer to the explanatory end. The comparator intervention will also be determined by the trial design. For a more pragmatic trial, the comparator will often be 'usual care' while for an explanatory trial, it may be a placebo intervention. If the comparator will be "usual care," it is important to recognize that usual care can have multiple meanings. For example, in some cases, it may be interpreted as being in

Table II
Strategies to minimize bias when therapist or participant blinding is not possible

- Blind the participants to the research hypothesis if possible and ethical, such as in a study comparing two different rehabilitation interventions that could each feasibly be of benefit to the participant.
- Ensure that treatment groups are treated equally as far as possible with respect to number of treatment sessions with the treating therapists, duration of treatment, quantity and quality of participant materials (e.g., written materials, equipment *etc.*). Detailed treatment protocols for each group will be required and information about how standardisation of therapy across groups was achieved and monitored must be recorded.
- Treating therapists must be educated and trained in the importance of treating participants from each treatment group equally and not to “favour” one group over another. Random auditing of therapist treatment sessions can provide assurance that groups are being treated as equally as possible by therapists.
- In cases where the same therapists are delivering the treatment interventions to both treatment groups, strategies should be put in place to prevent participants from one treatment group meeting participants from the other group and potentially discussing their treatment allocation (e.g., arranging appointments so that participants from each group are never treated at the same time or could meet each other in the waiting room)
- Researchers may choose to use separate pools of therapists to deliver each intervention arm (cluster designs).

accordance with clinical guidelines, while in others, it may simply mean management by a primary care physician. We recommend that when “usual care” is to be used as the comparator that authors provide detailed descriptions of what constitutes “usual care” treatment in their studies.

It is also important to consider the extent to which strategies to improve adherence to the interventions will be used and the nature of such strategies (e.g., the use of ‘booster’ sessions with a therapist to improve home exercise adherence or phone calls to remind participants to wear orthotics). In trials that are more explanatory, consideration of relevant strategies to maximize participant adherence to the intervention is paramount.

In all trials, a precise description of the intervention and the comparator, as well as any adherence strategies, needs to be provided so as to be reproducible by another team. Where individualized treatment is provided, a description of the procedures for tailoring the intervention to individual participants should be fully described¹¹. Descriptions may include: treatment setting and provider, number of health provider visits, group or individual sessions, duration of visits, frequency of visits; type of exercise, position of exercise, number of exercises, sets and repetitions, intensity of the program, degree of supervision, progression; model, manufacturer details and material for braces, orthotics or other aids and/or appliances; descriptions of fitting instructions and any instructions provided to participants regarding frequency of wear/wear of devices (e.g., shoe orthoses may be initially worn for 1 h per day and increased gradually over time until wearing them full-time); type of technique, patient position, grade and number of sets and repetitions for manual techniques. Some authors have used a goal-based rehabilitation program to tailor interventions to the individual patient¹². For example, the program may include pre-defined goals for joint motion, muscle function, and functional performance for patients to progress the rehabilitation activities from lower levels to higher levels of intensity¹². When using this approach, the operational definitions for achieving the goals must be described so that others can reproduce the approach.

It is also important to record the adherence to the intervention through the trial to assess the accuracy and generalizability of the results. The intensity to which this is done will be greater in explanatory compared to pragmatic trials. However, measuring participant adherence is not necessarily straightforward and the decision about the best method to use will need to take into account reliability and validity of the tool, participant-burden and resources. For example, self-reported measures of exercise adherence or splint, orthotic or brace usage may be recorded in a log book. Accelerometers or online registration of participation’s physical activity may be used to assess the amount of physical activity. It may also be possible for researchers to visually inspect devices (e.g., footwear or orthoses) at completion of the trial for degree of wear as another indicator of adherence. Developing proxy measures for adherence to exercise programs (similar to how blood sugar may be used as a proxy

measure for adherence to a dietary intervention in patients with diabetes as opposed to self-reported food intake) may be an interesting idea to explore. At this time there is little evidence to guide this approach but this is an area that would be worth further investigation. It is important to consider the extent to which monitoring of patient adherence actually influences adherence behavior itself. Publishing the treatment protocol separately prior to the main trial findings or as an appendix to the main trial paper is an excellent way to provide more detailed description of the treatment than is generally allowed in the main paper.

Unlike pharmacological therapies, the treatment effects are also likely to be influenced by the healthcare practitioner who performs or delivers the intervention. Given then that the healthcare practitioner is an integral part of the intervention, eligibility criteria and training of the practitioner should be considered and reported. In line with the CONSORT extension for non-pharmacological treatments¹¹, practitioner characteristics including gender, years of experience, previous research experience and qualifications need to be reported. In explanatory trials, a high level of practitioner skill and expertise in applying and monitoring the intervention is desirable to maximize treatment fidelity. Pragmatic trials will be more flexible in selecting practitioners and will generally include a range of skill levels. Training should be provided to practitioners and may take the form of study manuals, DVDs and/or face-to-face or on-line training programs. In some trials it may be necessary to formally ascertain that the practitioner has attained a certain level of skill following training before treating participants as part of the trial. Adherence of practitioners to the intervention throughout the trial should also be maximized, particularly in explanatory trials. This can be achieved through strategies such as holding regular meetings, booster training sessions, and providing treatment notes. Assessment of practitioner adherence to the protocol is important and can be achieved through methods such as audio/videotaping treatment sessions and auditing treatment notes. The number of practitioners chosen to deliver the intervention will vary depending on the trial requirements. Having a greater number of practitioners will increase the variability in treatment delivery but will enhance the external validity of the study.

Participation retention

Strategies for participation retention need to be described for all rehabilitation trials. In explanatory studies, participation retention strategies need to be maximized. For example, as mentioned earlier, in device trials, “run-in” methods (where participants are provided with a similar device to the one being studied (i.e., an alternate type of orthotic so as not to unblind participants) and only those who are sufficiently adherent to use of the device over an initial short time-frame are enrolled) might be incorporated to maximize intervention adherence and participant retention. In more pragmatic trials, participation retention strategies should be

enhanced but not to level that it might be an intervention itself. Method of communication and patient contact, visit reminders, transportation issues and other corresponding issues should be addressed and the reason for eventual discontinuation of participation given. In case of patient discontinuation in a trial, clinical assessment should be made or a preplanned and standardized telephone interview should be attempted. These strategies need to be included in the planned protocol as well as in the consent form.

Concomitant therapy

Symptomatic drug therapies are common for patients with symptomatic OA and may be used by participants in both non-operative and post-operative rehabilitation trials. The use of pharmacological treatment may interfere with study interventions and therefore affect outcome measures, but discontinuation might neither be ethical nor practical. In explanatory trials these concomitant therapies should be limited, but in more pragmatic trials continuing concomitant therapies could be desirable to enhance external validity. It is important to monitor the use of any other treatments that the patients would need or take during the trial and adjust for in the analysis to ensure that the effects of the rehabilitation interventions on symptoms or disease progression do not bias the outcome. The use of weekly or monthly web based communication systems or written log-books may be advisable to obtain the most reliable data.

Adverse events

Adverse events should be defined a priori and carefully recorded by the therapist and/or the participant and if appropriate reported to regulatory authorities in accordance with requirements in good clinical practice and according to the specific regulations and requirements in the different countries. Adverse events in rehabilitation trials are usually relatively minor, for example exacerbation of pain or swelling, and would not require report to regulatory authorities. A pain monitor scale where zero is “no pain” and ten is “worst considerable pain” should be used. Pain up to two is considered “safe” and up to five is considered “acceptable” as long as it is temporary¹³. Swelling should be monitored and the Stroke test could be used as an effusion grading scale.¹⁴

In instances where adverse events will need to be reported to regulatory/oversight boards, a determination of whether or not the event was related to study interventions or procedures will have to be determined. Data Safety and Monitoring Boards frequently require investigators to record all adverse events whether they believe the event is related to study interventions or not. This determination of whether the adverse event is study related should be made by an individual who is blinded to study group assignments to avoid any potential for bias in making this determination. Although the treating therapist may be the one to file an initial report, the therapist is not blinded to group assignment and should not make a final decision as to whether the event was study-related. The therapist can report the nature of the adverse event and the conditions under which it occurred without disclosing the treatment group assignment. The principal investigator or other designated blinded investigator could then determine if the adverse event was treatment-related without knowing which treatment group the participant belongs to and hence avoid (or minimize) the potential bias.

Protocol violations

Protocol violations in rehabilitation studies can include violations in randomization, intervention, and outcome procedures. It is

important to record protocol violations as they occur for all participants in a study as the nature and frequency of these violations could have implications for decisions regarding participant termination, or in studies where intention-to-treat approaches are employed, determining if protocol deviations need to be controlled for in the final data analysis. Likewise, it would also be important to record the number of participant dropouts between groups and determine if there were group differences on this factor that may bias any group effects.

Violations in randomization may include randomization of participants who do not meet the study inclusion criteria or when participants are mistakenly given the treatment approach contrary to the one in which they were randomly assigned. Violations in intervention procedures may include the administration of improper treatment dosage (intensity, frequency, duration) or administration of co-interventions not specified as approved co-interventions in the study protocol. In some cases, participants may cross-over to the other intervention group during the course of the study. This might occur in study designs where one group receives an experimental intervention combined with standard treatment and is compared to a group receiving only the standard treatment. A participant assigned to the experimental treatment group may refuse to continue to receive the experimental intervention and thus only receives the standard treatment for the remainder of the study. Outcome procedure protocol violations could include failure to administer a specific outcome measure at a given follow-up time point. Outcome procedure protocol violations might also include obtaining measurements out of the appropriate time window for the follow-up period or failing to obtain any outcome measures for a given follow-up period.

A report should be completed for each protocol violation that occurs for each participant in the study. The report should be kept in the participant's study record. When completing a report the investigators should include the date the protocol violation occurred, the name of the individual completing the report, and a description of the nature of the protocol violation. Protocol violations related to outcome measures should include the specific measure or measures that are missing and the follow-up period associated with the violation. Any explanations for why the protocol violation occurred should also be included in the report.

In the final publication, it would not be necessary to report each and every protocol violation that occurred over the course of the study. However, any protocol violations that resulted in the termination of participant involvement in the study should be reported. Likewise, if any category of protocol violation was identified as a potential co-variate that needed to be controlled for in the final analysis, this should also be reported. In studies where intention-to-treat approaches to data analysis are employed, the data from participants who incurred protocol violations over the course of the study would continue to be included in the final analysis.

Outcome measures

Primary outcomes

Explanatory or efficacy clinical trials should have a clearly defined primary outcome variable. This outcome variable will depend on the nature of the study, but should be a direct and immediate consequence of the intervention⁵. For pragmatic trials, the primary outcome should be clinically meaningful, but may not be the direct result of the intervention. For example, in an explanatory trial evaluating a stretching intervention, the primary outcome should be range of motion. For a pragmatic trial of a similar nature, the primary outcome may be a measure of functional performance or a patient-reported outcome measure.

An alternative approach is to use several, pre-defined primary outcome measures. This approach may require adjustments to the statistical significance to account for multiple analyses. For all trials, the outcome measure must be operationally defined if there is no universal standard for quantification.

Secondary outcomes

Inclusion of one or more secondary outcomes is recommended in rehabilitation studies. Acquisition of secondary outcomes should not interfere with the primary outcome measure and the study should be appropriately powered if the intention is to evaluate the effect of the intervention on these outcomes.

Outcome measures of OA

Outcome measures should consist of variables that have been reported to be reliable, valid and responsive within the clinical population included in the study. Table III provides a list of outcome measures that may be used in OA rehabilitation trials.

Pain. The magnitude of pain in the joint of interest should be quantified. Pain can be quantified using a Likert scale of severity in which the options include “None”, “Mild”, “Moderate”, “Severe”, or “Very Severe.” Other options include a 100 mm Visual Analog Score (VAS) in which the end-points are anchored with “No Pain” and “Worst Pain Imaginable.” Alternatively, an 11-point Likert scale can be used with anchors of “0 – No Pain” and “10 – Worst Pain Imaginable.” The specific type of pain (e.g., pain at rest, pain with motion, etc.) and time-frame of pain recall (e.g., maximal pain over the last week, current pain, etc.) should be documented. The time allowed between the episode of pain and the recall of pain should be minimized, as longer recall times reduce the ability to accurately recall pain severity. Optimal time for pain recall is less than 48 h¹⁵. Activity-specific questions of pain may be appropriate and can be measured using a Likert score or VAS (e.g., pain during weight-bearing or pain during stair-climbing). Validated instruments with pain subscales can also be used (e.g., Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) pain subscale¹⁶, Knee Injury and Osteoarthritis Outcome Scale (KOOS) pain subscale¹⁷, Hip Disability and Osteoarthritis Outcome Score (HOOS) pain subscale¹⁸, or the Boston University Osteoarthritis Computer Adapted Test (BU OA-CAT) Functional Pain Scale¹⁹). Valid and responsive pain scales that capture both constant and intermittent pain, such as the Intermittent and Constant Pain Score (ICOAP) may also be of value in trials evaluating symptom modifying interventions²⁰. The reader is referred to the articles on knee, hip, and hand OA trials in this special edition for detailed descriptions of potential symptom measures.

Patient-reported physical function. Valid, reliable and responsive self-report of physical function questionnaires or subscales are recommended for studies of OA. These questionnaires may be disease-specific (e.g., WOMAC or Arthritis Impact Measurement Scale (AIMS)²¹) or joint- and disease-specific (Knee Injury and Osteoarthritis Outcome Score (KOOS)²² and HOOS²³ Activities of Daily Living and Sport and Recreation and Physical Function subscales, Knee or Hip Outcome Scale Activities of Daily Living Scale (K/HOS-ADLS)^{24,25}, Australian/Canadian Hand Osteoarthritis Index (AUSCAN)²⁶, or Functional Index of HOA (FIHOA)^{27,28}.) General or joint-specific functional scales are appropriate if they have been validated in similar OA populations (e.g., Lower Extremity Functional Scale (LEFS)).²⁹

Patient global assessment. The patient's perception of his or her overall functional ability should be assessed using a single Likert or VAS metric. A single optimal method has not been established,

Table III

Categories of outcome measures in OA rehabilitation trials

Recommended Core Set of Measures
<ul style="list-style-type: none"> • Pain • Patient-reported outcome measures • Performance-based measures • Patient global assessment
Additional Measures Strongly Recommended
<ul style="list-style-type: none"> • Health-related quality of life • Ratio/number of responders
Optional Measures
<ul style="list-style-type: none"> • Signs of inflammation (redness, swelling/joint effusion, increased skin temperature) • Self reported stiffness • Range of motion • Biomechanical parameters derived from gait analysis • Laxity • Muscle strength • Proprioception • Self-reported instability (buckling or giving way, etc.) • Body weight/BMI • Time to surgery • Medication consumption • Physical activity • Psychological measures • Biomarkers • Imaging of structural change • Cost-effectiveness • Adherence • Acceptability/comfort (e.g., for a study of orthotics or devices)

although a standard question should be asked at all evaluation points in the trial. For example, “How would you rate your current level of function during your usual activities of daily living?” can be quantified on a 100 mm VAS with end-points anchored as “Inability to perform any daily activities” and “No problem with any daily activity” to capture a participant's overall global rating. Single questions related to global rating of change may also be used to capture the patient's perception of change overall or pain from baseline to follow-up examination. If using the global rating of change, it should be clear to the patient the time frame over which you are asking about the change in condition. For example the query might include something like “based on how you were doing at the start of your treatment, how much improvement do you feel you have made up to now,” in order to let the patient know you want them to refer to change from baseline to the present. It has also been suggested to provide a memory marker (e.g., asking the patient some question about their condition at baseline, then reminding them of this statement at the follow-up) to the patient to assist in improving the patient's recall of their baseline condition.³⁰

Performance-based measures. Performance-based and self-reported functional assessments evaluate different domains of disability and are associated with different underlying impairments in patients with OA³¹. Therefore, performance-based tests are recommended as outcomes in rehabilitation clinical trials. Performance-based outcomes may include time to complete a task, the distance walked in a specified time, or the number of successful attempts in a given time to complete a functional movement. The OARSI initiative to develop a recommended set of physical performance measures for hip and knee OA identified a minimal core set of measures to evaluate functional performance of sit-to-stand, short walking distances and stair negotiation tasks. These recommended variables include the 30 s chair stand test, the 4 × 10 m fast-paced walk test and a timed stair task, respectively³². There currently is not an established minimal core set of performance-based measures for hand OA, however some options may include

grip strength, pinch grip, the Arthritis Hand Function Test, and Moberg's Pick-up Test.^{33–35}

Generic health status measures. Measurements of generic health status are strongly recommended for inclusion in rehabilitation trials for persons with OA when the duration of the trial is at least 6 months³⁶. The Short-Form 36 Health Survey (SF-36)³⁷, and the EuroQOL³⁸ are examples of health status measures that have been used and validated in patients with OA. Health status measures should be included if cost-effectiveness is used as a primary or secondary outcome.

Clinical joint examination. Joint-specific impairments such as joint circumference, effusion, range of motion, strength, laxity and proprioception may be important outcomes in explanatory trials. The utility of these measures in pragmatic clinical trials is dependent on the nature of the intervention and relation of these impairments to down-stream functional performance. Joint stiffness has been recommended to be assessed using patient-reported measures (e.g., items on WOMAC, KOOS or HOOS³⁶, but self-reported stiffness measures may not be as reliable or valid as self-reported pain measures. Joint alignment is associated with OA disease progression and there is some evidence that joint mal-alignment can affect the outcome of exercise on pain responses³⁹. Therefore, measurement of alignment may be important in rehabilitation trials. There are several methods for measuring joint alignment and the reader is referred to the other articles on imaging knee and hand for discussion on these methods.

Body weight and BMI. Body weight and BMI should be assessed at baseline and at follow-up evaluations for clinical trials that target symptom-modifying interventions. Change in body weight or BMI may be used as an outcome measure or as a covariate, when appropriate, in statistical analyses.

Time to surgery. Symptom-modifying treatments that attempt to eliminate or delay the need for surgical procedures (e.g., high-tibial osteotomy or joint arthroplasty) can be assessed using time-to-event analysis. However, the decision to undergo surgery is influenced by many environmental, health-related and psycho-social factors (e.g., ability to pay for procedure, social support following surgery, preclusive co-morbidities). In addition, there may be differences in health systems within and between countries that might influence the generalizability of the study findings.

These confounding factors must be recorded on an individual level or accounted for in the inclusion/exclusion criteria when time to surgery is used as a primary outcome.

Time to return to activity. Some rehabilitation trials, particularly post-operative rehabilitation trials, may consider time to return to work, sport, or recreational activity as an important outcome measure. These can also be assessed using time-to-event analysis. Similar to using time to surgery as an outcome, the decision to return to work, sport, or recreational activity may be influenced by environmental, health-related, and psycho-social factors. Again, these factors may need to be recorded on an individual level or accounted for in the inclusion/exclusion criteria when time to activity is used as a primary outcome.

Medication consumption. The use of rehabilitation interventions to eliminate or reduce the need for pharmacological management of symptoms is common. The type, dose, method of administration and frequency of analgesic, NSAID or symptom-modifying medication use should be recorded.

Physical activity and participation. The goals of many rehabilitation interventions are to improve physical activity and, ultimately, increase participation. If the outcome is to directly measure activity intensity or duration, investigators should use accelerometer-based measures of activity. In larger trials, the costs and time associated with accelerometer-based measures of physical activity may be prohibitive, in which case self-reported scales that have been validated in the target OA population may be used. These instruments include the Lower-Extremity Activity Scale, University of California Los Angeles Physical Activity Scale, The Physical Activity Scale for the Elderly, and the Human Activity Profile, among others^{40–42}. The choice of accelerometer and/or specific self-reported physical activity measures should be based on the individual study goals as the reliability and validity of these measures are dependent on aims of the clinical trial⁴¹. Furthermore, we need to acknowledge that the validity of some of these physical activity questionnaires is limited.⁴³

Psychological measures. There is a reciprocal interrelation between depression and musculoskeletal pain⁴⁴. Given the influence of psychological status on the magnitude of pain and the efficacy of rehabilitation interventions in this population, psychological measures that capture self-efficacy, anxiety, depression, coping strategies and pain catastrophizing may be important metrics to include as outcome measures or as covariates for statistical analyses. Examples include the Arthritis Self-Efficacy Scale⁴⁵, Beck Anxiety Index⁴⁶, Centers for Epidemiological Studies Depression Scale⁴⁷, Coping Strategies Questionnaire⁴⁸, and Pain Catastrophizing Scale⁴⁹, respectively.

Biomarkers and imaging. Although the aims of most symptom-modifying interventions are to reduce pain and improve function, some interventions may concomitantly delay joint deterioration or impact overall joint structure and metabolism (e.g., biomechanical, bracing or orthoses, weight loss or activity interventions). Biomarkers of structural OA progression or imaging of joint structure may be appropriate for interventions that may have an associated positive or negative structural or metabolic effect. Readers are referred to the recommendations established by the Imaging and Biochemical markers sub-groups, respectively.

Cost-effectiveness. We recommend that incremental cost-effectiveness and/or incremental cost-utility ratios (ICURs) be used as cost-effectiveness outcome measures for rehabilitation trials that aim to evaluate cost-effectiveness of intervention approaches⁵⁰. The incremental cost-effectiveness ratio (ICER) is an estimate of differences in costs of healthcare in relation to differences in the health outcomes obtained between groups, and is calculated by dividing the difference in healthcare costs between groups by the difference in the health outcome measure between groups⁵¹. The ICUR is an estimate of differences in costs of healthcare in relation to societal value of the resulting health outcomes (utility)⁵¹. Utility is usually represented by quality-adjusted life years (QALY). The ICUR is calculated by dividing the difference in healthcare costs between groups by the difference in QALYs between groups over the study follow-up period.

When using ICERs or ICURs as outcome measures, the methods for estimating costs, health outcomes, and health outcomes utility should be described. The estimation of cost should reflect direct and indirect costs for managing the participants' care. Direct medical costs can include items such as physician or other healthcare professional office visits, medical/surgical tests and procedures, medications, devices, equipment, etc. Non-medical direct costs are costs to the individual that may result from seeking medical attention or following through with medical

interventions and include items such as transportation, child care, home care, *etc.* Indirect costs may include items such as lost wages to the patient or family members, parking fees, home renovations, *etc.* Self-report instruments are commonly used and are acceptable for obtaining information to estimate direct and indirect healthcare costs.⁵²

Response criteria

In addition to evaluating differences between groups in the primary and secondary outcomes, responder rates of clinically meaningful differences can be quantified for the study population. The OARSI Standing Committee for Clinical Trials Response Criteria Initiative has identified responder criteria based on absolute and relative changes in pain, function and global assessment domains⁵³. It should be noted that these criteria are primarily derived from pharmacological trials and implementation into rehabilitation trials has not been fully evaluated. Variable-specific clinically meaningful differences or changes should be considered when evaluating response criteria. Investigators are cautioned however, to take care in interpreting minimally important differences. As stated by King;

“There is no universal MID, despite the appeal of the notion. Indeed, for a particular patient-reported outcome instrument or scale, the MID is not an immutable characteristic, but may vary by population and context. At both the group and individual level, the MID may depend on the clinical context and decision at hand, the baseline from which the patient starts, and whether they are improving or deteriorating. Specific estimates of MIDs should therefore not be overinterpreted.”

Benchmarking outcomes to established values

Outcomes from all arms of the clinical trial may be compared to previously established age- or condition-matched normative values if available. Benchmarking outcomes from a rehabilitation clinical trial reduces the possibility of identifying a superior treatment that fails to meet or exceed known outcomes.

Statistical considerations

Details for addressing the description of statistical analysis procedures and reporting of results can be obtained in the article concerning statistical considerations that accompanies this special issue. Here we provide some basic information that should be included in describing data analysis and reporting of results for rehabilitation clinical trials.

Sample size estimates should be described for primary outcome variables. The description should include the statistical test for which the sample size was estimated (e.g., two way analysis of variance, *etc.*) as well as the effect size, level of statistical significance, level of statistical power, and the expected dropout rate that were used to calculate the sample size estimate. The sources used to justify the effect size and expected drop out rates should be provided in the description so that the reader has enough information to reproduce the sample size estimates presented by the authors.

Methods used for randomization and allocation concealment should be described. The description should include the conditions under which stratification and/or block allocation methods were employed if applicable. If computer generated methods or online services for randomization were used this should be indicated. If online services are publicly available they should be cited in the manuscript. The method for allocation concealment (computer generated electronic notification, sealed envelope, *etc.*) should be described.

A description of all statistical tests used in the study should be provided. Primary focus should be given to the primary outcomes of the study and all additional analyses should be described as secondary analyses. It may be helpful to present the description of statistical tests as it relates to each aim in the study (e.g., “to compare mean differences between groups over time on pain scores, a two way repeated measures analysis of variance was performed.”) There should be a description provided of how data was screened to determine they met the assumptions for the tests being used in the analysis (e.g., if parametric tests were used were the data screened for normality, homogeneity of variance, *etc.*) The level of significance for each statistical test should be stated and there should be a description of methods used to adjust for multiple comparisons if applicable. If interim analyses were performed the method to adjust for this should be described. Descriptions of any adjustments for confounders/covariates should also be provided.

A statement should be provided indicating whether the analysis was performed using intention-to-treat methods or as-treated analyses (as may be the case in efficacy trials) were employed. If intention-to-treat methods are used, then the methods for dealing with missing data (e.g., multiple imputation, last score forward, *etc.*) should also be described. It may be helpful to characterize the missing data as missing completely at random, missing at random, or not missing at random.

There are many methods for reporting the results of clinical trials that often include the use of data tables, charts and figures. Regardless of the method an author may select, there are certain elements that we recommend be included in reporting of results for rehabilitation trials. These elements would facilitate the use of the study data in subsequent systematic reviews and meta-analyses. The means, standard deviations, and 95% confidence intervals (CIs) should be provided for the primary outcome for each study group. Authors should also report mean differences and the standard deviations and 95% CIs of the differences between groups for the primary outcome. In cases where loss to follow-up may have occurred, the sample size for each group in which the analyses were performed should also be provided.

Author contributions

All authors contributed to the conception and design of the guidelines and recommendations, drafting the article and revising it critically for important intellectual content, and provided final approval of the version to be submitted. Dr Fitzgerald (kfitzger@pitt.edu) takes responsibility for the integrity of the work as a whole, from inception to finished article.

Other contributors

Not applicable.

Statement of role of funding source in publication

Not applicable.

Conflict of interest

None of the authors have conflicts of interest to declare.

Acknowledgment

OARSI gratefully acknowledges support to defer in part the cost of printing of the Clinical Trial Recommendations from Abbvie, BioClinica, Boston Imaging Core Lab, and Flexion. The funding sources for printing had no role in the outcome of this manuscript.

References

- Altman R, Brandt K, Hochberg M, Moskowitz R, Bellamy N, Bloch DA, et al. Design and conduct of clinical trials in patients with osteoarthritis: recommendations from a task force of the Osteoarthritis Research Society. *Osteoarthritis Cartilage* 1996;4:217–43.
- Altman R, Asch E, Bloch D, Bole G, Brandt KD, Christy W, et al. Development of criteria for the classification and reporting of osteoarthritis. *Arthritis Rheum* 1986;29:1039–49.
- Altman R, Alarcon G, Appelrouth D, Bloch D, Borenstein D, Brandt K, et al. The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hand. *Arthritis Rheum* 1990;33:1601–10.
- Altman R, Alarcon G, Appelrouth D, Bloch D, Borenstein D, Brandt K, et al. The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hip. *Arthritis Rheum* 1991;34:505–14.
- Thorpe K, Zwarenstein M, Oxman A, Treweek S, Furberg CD, Altman DG, et al. A pragmatic-explanatory continuum indicator summary (PRECIS): a tool to help trial designers. *J Clin Epidemiol* 2009;62:464–75.
- Hutchings A, Calloway M, Choy E, Hooper M, Hunter DJ, Jordan JM, et al. The Longitudinal Examination of Arthritis Pain (LEAP) study: relationships between weekly fluctuations in patient-rated joint pain and other health outcomes. *J Rheumatol* 2007;34:2291–300.
- Bellamy N, Sothorn RB, Campbell J. Rhythmic variations in pain perception in osteoarthritis of the knee. *J Rheumatol* 1990;17:364–72.
- Paterson C, Dieppe P. Characteristic and incidental (placebo) effects in complex interventions such as acupuncture. *Br Med J* 2005;330:1202–5.
- Wood L, Egger M, Gluud L, Schulz KF, Juni P, Altman DG, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *Br Med J* 2008;15:601–5.
- Zhang W, Nuki G, Moskowitz R, Abramson S, Altman DG, Arden NK, et al. OARSI recommendations for the management of hip and knee osteoarthritis: Part III: changes in evidence following systematic cumulative update of research published through January 2009. *Osteoarthritis Cartilage* 2010;18:476–99.
- Boutron I, Moher D, Altman D, Schulz K, Ravau P, Group C. Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: explanation and elaboration. *Ann Intern Med* 2008;148:295–309.
- Frobell RB, Roos EM, Roos HP, Ranstom J, Lohmander L. A randomized trial of treatment for acute anterior cruciate ligament tears. *N Engl J Med* 2010;363:331–42.
- Ageberg E, Link A, Roos E. Feasibility of neuromuscular training in patients with severe hip or knee OA: the individualized goal-based NEMEXTJR training program. *BMC Musculoskelet Disord* 2010;11:126.
- Sturgill L, Snyder-Mackler L, Manal T, Axe M. Interrater reliability of a clinical scale to assess knee joint effusion. *J Orthop Sports Phys Ther* 2009;39:845–9.
- Perrot S, Marty M, Legout V, Moysse D, Henrotin Y, Rozenburg S. Ecological or recalled assessments in chronic musculoskeletal pain? A comparative study of prospective and recalled pain assessments in low back pain and lower limb painful osteoarthritis. *Pain Med* 2011;12:427–36.
- Bellamy N, Watson-Buchanan W, Goldsmith C, Campbell J. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *J Rheumatol* 1988;15:1833–40.
- Roos E, Lohmander L. The Knee Injury and Osteoarthritis Outcome Score (KOOS): from joint injury to osteoarthritis. *Health Qual Life Outcomes* 2003;1:64–71.
- Nilsdotter A, Lohmander L, Klassbo M, Roos E. Hip disability and osteoarthritis outcome score (HOOS)-validity and responsiveness in total hip replacement. *BMC Musculoskelet Disord* 2003;4.
- Jette A, McDonough C, Ni P, Haley SM, Hambleton RK, Olarsch S, et al. A functional difficulty and functional pain instrument for hip and knee osteoarthritis. *Arthritis Res Ther* 2009;R107.
- Hawker G, Davis A, French M, Cibere J, Jordan JM, March L, et al. Development and preliminary psychometric testing of a new OA pain measure—an OARSI/OMERACT initiative. *Osteoarthritis Cartilage* 2008;16:409–14.
- Meenan R, Gertman P, Mason J. Measuring health status in arthritis. The arthritis impact measurement scales. *Arthritis Rheum* 1980;23:146–52.
- Roos E, Roos H, Lohmander L, Ekdahl C, Beynon B. Knee Injury and Osteoarthritis Outcome Score (KOOS)—development of a self-administered outcome measure. *J Orthop Sports Phys Ther* 1998;28:88–96.
- Klassbo M, Larsson E, Mannevik E. Hip disability and osteoarthritis outcome score. An extension of the Western Ontario and McMaster Universities Osteoarthritis Index. *Scand J Rheumatol* 2003;32:46–51.
- Irrgang J, Snyder-Mackler L, Wainner R, Fu F, Harner C. Development of a patient-reported measure of function of the knee. *J Bone Joint Surg Am* 1998;80-A:1132–45.
- Naal F, Impellizzeri F, von Eisenhart-Rothe R, Mannion A, Leunig M. Reproducibility, validity, and responsiveness of the hip outcome score in patients with end-stage hip osteoarthritis. *Arthritis Care Res* 2012;64:1770–5.
- Bellamy N, Campbell J, Haraoui B, Buchbinder R, Hobby K, Roth JH, et al. Dimensionality and clinical importance of pain and disability in hand osteoarthritis: development of the Australian/Canadian (AUSCAN) Osteoarthritis Hand Index. *Osteoarthritis Cartilage* 2002;10:855–62.
- Dreiser R, Maheu E, Guillou G, Caspard H, Grouin J. Validation of an algofunctional index for osteoarthritis of the hand. *Rev Rhum Engl Ed* 1995;62(6 Suppl 1):43S–53S.
- Moe R, Garratt A, Slatkowsky-Christensen B, Maheu E, Mowinckel P, Kvien TK, et al. Concurrent evaluation of data quality, reliability and validity of the Australian/Canadian Osteoarthritis Hand Index and the Functional Index for Hand Osteoarthritis. *Rheumatology* 2010;49:2327–36.
- Pua Y-H, Cowan S, Wrigley T, Bennell K. The lower extremity functional scale could be an alternative to the Western Ontario and McMaster Universities Osteoarthritis Index physical function scale. *J Clin Epidemiol* 2009;62:1103–11.
- Wyrwich K, Wolinsky F. Identifying meaningful intra-individual change standards for health-related quality of life measures. *J Eval Clin Pract* 2000;6:39–49.
- Stratford P, Kennedy D. Performance measures were necessary to obtain a complete picture of osteoarthritic patients. *J Clin Epidemiol* 2006;59:160–7.
- Dobson F, Hinman R, Roos E, Abbott JH, Stratford P, Davis AM, et al. OARSI recommended performance-based tests to assess physical function in people diagnosed with hip or knee osteoarthritis. *Osteoarthritis Cartilage* 2013;21:1042–52.
- Maheu D, Altman R, Bloch D, Doherty M, Hochberg M, Mannoni A, et al. Design and conduct of clinical trials in patients with osteoarthritis of the hand: recommendations from a task force of the Osteoarthritis Research Society International. *Osteoarthritis Cartilage* 2006;14:303–22.

34. Backman C, Mackie H. Arthritis hand function test: inter-rater reliability among self-trained raters. *Arthritis Care Res* 1995;8:10–5.
35. Ng C, Ho D, Chow S. The Moberg pickup test: results of testing with a standard protocol. *J Hand Ther* 1999;12:309–12.
36. 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.
37. Ware Jr J, Kosinski M, Bayliss M, McHorney C, Rogers W, Raczek A. Comparison of methods for the scoring and statistical analysis of SF-36 health profile and summary measures: summary of results from the Medical Outcomes Study. *Med Care* 1995;33(Suppl 1):AS264–79.
38. Fransen M, Edmonds J. Reliability and validity of the EuroQol in patients with osteoarthritis of the knee. *Rheumatol* 1999;38:807–13.
39. Lim B-W, Hinman R, Wrigley T, Sharma L, Bennell K. Does knee malalignment mediate the effects of quadriceps strengthening on knee adduction moment, pain, and function in medial knee osteoarthritis? A randomized controlled trial. *Arthritis Care Res* 2008;59:943–51.
40. Bennell K, Hinman R, Crossley K, Metcalf BR, Buchbinder R, Green S, et al. Is the Human Activity Profile a useful measure in people with knee osteoarthritis? *J Rehabil Res Dev* 2004;41:621–30.
41. Terwee C, Bouwmeester W, van Elstrand S, de Vet H, Dekker J. Instruments to assess physical activity in patients with osteoarthritis of the hip or knee: a systematic review of measurement properties. *Osteoarthritis Cartilage* 2011;19:620–33.
42. Martin KARW, Miller ME, James MK, Ettinger WH, Messier SP. Validation of the PASE in older adults with knee pain and physical disability. *Med Sci Sports Exerc* 1999;31:627–33.
43. Svege I, Kolle E, Risberg M. Reliability and validity of the Physical Activity Scale for the Elderly (PASE) in patients with hip osteoarthritis. *BMC Musculoskelet Disord* 2012;13:26.
44. Kroenke K, Wu J, Bair M, Krebs E, Damush T, Tu W. Reciprocal relationship between pain and depression: a 12-month longitudinal analysis in primary care. *J Pain* 2011;12:964–73.
45. Lorig K, Chastain R, Ung E, Shoor S, Holman H. Development and evaluation of a scale to measure perceived self-efficacy in people with arthritis. *Arthritis Rheum* 1989;32:37–44.
46. Beck AT, Epstein N, Brown G, Steer RA. An inventory for measuring clinical anxiety: psychometric properties. *J Consult Clin Psych* 1988;56:893–7.
47. Radloff LS. The CES-D scale: a self-report depression scale for research in a general population. *Appl Psychol Meas* 1977;1:385–401.
48. Rosenstiel A, Keefe F. The use of coping strategies in chronic low back pain patients: relationship to patient characteristics and current adjustment. *Pain* 1983;17:33–44.
49. Sullivan M, Bishop S, Pivik J. The pain catastrophizing scale: development and validation. *Psychol Assess* 1995;7:524–32.
50. Pinto D, Robertson M, Hansen P, Abbott JH. Cost-effectiveness of nonpharmacological, nonsurgical interventions for hip and/or knee osteoarthritis: systematic review. *Value Health* 2012;15:1–12.
51. Homick J, Suarez-Almazor M. An economic approach to healthcare. *Best Pract Res Clin Rheumatol* 2004;18:203–18.
52. Pinto D, Robertson M, Hansen P, Abbott JH. Good agreement between questionnaire and administrative databases for health care use and costs in patients with osteoarthritis. *BMC Med Res Methodol* 2011;11:45.
53. Pham T, van der Heijde D, Altman RD, Anderson JJ, Bellamy N, Hochberg M, et al. OMERACT-OARSI initiative: Osteoarthritis Research Society International set of responder criteria for osteoarthritis clinical trials revisited. *Osteoarthritis Cartilage* 2004;12:389–99.