eStudy protocol
Efficacy of pharmacological treatments in long-term randomized controlled trials in knee osteoarthritis: a systematic review and meta-analysis

Protocol date: 15 September 2015, amended on 8 June 2017.

Introduction

Background and rationale
Knee osteoarthritis (OA) is one of the most common causes of pain, functional loss and disability, that substantially reduces subject’s autonomy in daily activities. In the next future, the global burden of knee OA will become a major health problem as a consequence of aging process, systemic diseases, genetic predisposition and environmental factors. Prevalence of the disease varies between different countries, populations and OA definitions (clinical or radiological). It has been recognized that incidence of knee degeneration is more frequent among elderly and female subjects, and increases in case of obesity, varus knee alignment and presence of OA in multiple joints.

Knee OA is a complex chronic disease involving not only cartilage structure, but also ligament, meniscus and periarticular muscles. Pain and function loss are the typical symptoms usually related to the worsening of the pathology: together with signs of crepitus, restricted movement and bony enlargement, they are the key elements to define a confident diagnosis. Moreover, the measurement of radiological joint space narrowing is another method currently used to monitor progression of the disease, especially for patients treated with disease-modifying drugs or Symptomatic Slow Acting Drugs for OA (SYSADOAs).

Current treatments for knee OA include pharmacological agents alleviating the symptomatology and others able to modify the disease process. The first type of agents include all pain relief drugs and all substances reducing joint inflammation (e.g., steroids), while the second type of pharmacological treatments consists of agents with putative disease-modifying properties such as glucosamine, chondroitin sulfate, vitamins, hyaluronic acid, etc.

Several studies have been conducted with the purpose to identify the most effective pharmacological treatment. However, although OA is a chronic and progressive disease, the vast majority of pharmacological agents are studied for their short-term effects on symptoms.

When a patient’s clinical condition is not suitable for surgical management, pharmacological therapy is the only one solution: for this reason, research should be focused on the comparison between different treatments, with the specific aim of identifying the best pharmacological choice able to improve the quality of life of the subjects and to reduce pathology progression over a long-term observation period. The present meta-analysis aims to offer a clear answer to those questions related to the treatment of knee OA, by evaluating long-term effects of different pharmacological therapies for knee OA on both symptoms and radiological signs.

Objectives
The present study evaluates the effectiveness (over a long-term period) of currently available pharmacological treatments for knee OA, in terms of pain reduction, preservation of functionality and of radiological joint space width.

Methodology

Eligibility criteria
According to criteria proposed by the PRISMA Extension Statement for Reporting of Systematic Reviews Incorporating Network Meta-analyses of Health Care Interventions, the selection of studies will be based on the specific definition of PICOS items: Participants, Interventions, Comparators, Outcomes and Study design.

 Characteristics of participants
Studies of interest will include adult participants, at least 18 years old, of both sexes and with a diagnosis of primary knee OA, possibly according to the criteria proposed by the American College of Rheumatology or similar. The presence of concomitant diseases or conditions, e.g., obesity or diabetes will not represent a reason for study exclusion. Moreover, studies involving mixed samples of patients with OA of knee and/or hip may also be included, if retrieved and appropriate for other criteria.

 Characteristics of interventions
The study will focus on most commonly prescribed pharmacological treatments for knee OA, including analgesics Nonsteroidal Anti-Inflammatory Drugs (NSAIDs), Hyaluronic Acid, Vitamin supplementation, corticosteroids, Symptomatic Slow Acting Drugs for Osteoarthritis (SYSADOA) or putative disease modifiers, and others. Pharmacological interventions have been identified following recent guidelines on knee OA treatment and recent clinical trials. Both clinical trials with monotherapy and clinical trials with combination therapy will be evaluated as eligible for final analysis if they include pharmacological treatments of interest.
Characteristics of comparators

Studies will be included if comparing a pharmacological treatment for knee OA with an active control (eg, another therapy for OA), or a control, including the administration of a placebo treatment, or no treatment.

Characteristics of outcomes

Authors will include in the analysis all articles reporting data for at least one of the following outcomes: pain, function and radiological joint space width. When a single outcome on pain or function is measured by different scale systems in the same article, the prioritized list of patient-reported outcomes proposed by Juhl will be considered. Eligible outcomes will be measured at 12-months follow-up or longer. The full text of the studies will be analyzed by two reviewers, independently, and articles with sufficient information about at least one outcome of interest will be included. For each outcome, data (i.e. baseline and follow-up values and/or mean change from baseline) will be extracted in case they are expressed as numerical and/or graphical information. In the latter case, numbers will be extracted from graphs using the procedure (adapted) from Guyot. Only studies with available data will be included in the analysis: no additional information on unpublished data will be requested to the authors of papers.

Characteristics of study design

Papers will be included if they used a randomized, controlled, blinded or non-blinded design. Case series and case reports will be excluded. Authors will select studies with no restriction criteria in terms of publication: both published and unpublished trials will be included, the latter including e.g. abstracts, conference proceedings and posters with available data. Moreover, no restriction will be applied in terms of language and type of setting. Studies comparing interventions of interest and reported extractable data for at least one measure of pain, function or joint space, will be included. Intention-to-treat data will be used whenever available.

Primary and secondary endpoints

Our primary endpoint will be to evaluate effectiveness of pharmacological treatments on knee pain. Moreover, we will consider function loss and maintenance of radiological joint space width as secondary endpoints.

Information sources and search procedure

The search procedure will be implemented consistently with the following criteria.

Electronic source and search strategy

Studies will be retrieved through a systematic review of current available literature on pharmacological treatment for knee OA. We will search the online databases PubMed, Scopus, Embase, Web of Science and Cochrane Central Register of Controlled Trials and for publication, from inception to August 31st 2017. The search strategy is provided separately and will be attached to the final report.

Hand-searching

Documentation from conference proceedings, abstracts and poster sessions will be hand searched consulting online sources of major international association involved in OA research: European League Against Rheumatism (EULAR), Osteoarthritis Research Society International (OARSI), American Academy of Orthopedic Surgeons (AAOS) and American College of Rheumatology (ACR). Moreover, references retrieved from meta-analyses and review articles will be hand-searched and analyzed and included if consistent with the selection criteria of the studies.

Study selection

Study identification will be performed by two reviewers, independently, and each potential discrepancy will be discussed and solved through consensus and independent expert consultation. The process of selection will be implemented consistently with the screening phases proposed by the PRISMA statement, and will be managed using the EndNote program, a software tool for publishing and managing bibliographies, citations and references. Articles will be screened by title, abstract and full text analysis considering selection criteria defined in this protocol. At the end of each stage, the reviewers will discuss reasons for inclusion and/or exclusion, and when consensus will not be reached, a further judgment will be requested to an independent expert in order to obtain a shared agreement on those studies selected for the meta-analysis.

Assessment of risk of bias

Methodological quality of selected studies will be evaluated using both the Cochrane risk of bias tool and the Jadad score. This double check method will reduce the probability of an incorrect or inaccurate judgment risk of bias: due to the extensive number of trials on knee OA, quality evaluation process is fundamental to achieve a reliable result.
Each article will be evaluated independently in a blind method by four researchers: two of them will use the Cochrane risk of bias tool, while the other two will complete the Jadad score. The latter is a standard scale scoring of five items, i.e. if the study was described as randomised and/or double-blind (0=no, +1=yes), whether the description of randomisation and/or double-blinding were reported and appropriate (0=no, +1=yes appropriate, -1=eyes and inappropriate), similarly to the description of withdrawals (0=no description, +1=yes and adequate imputation of missing data with intention-to-treat analysis, -1=yes and inadequate or no imputation of missing data). The maximum possible score is 5. The Cochrane risk of bias tool considers characteristics of the following items: sequence generation, allocation concealment, blinding of participants, study personnel and outcome assessors, incomplete outcome data, selective outcome reporting and other potential sources of bias. At the end of quality assessment, a consensus on final evaluation will be reached; any disagreements will be resolved by the judgement of another author.

Data extraction

Two reviewers will extract data independently from original papers reporting for each article the following information: study design, characteristics of population (age, sex and BMI), sample size, pharmacological treatment, time of follow-up, route of administration, type of placebo or other control, measurement tool, scores at baseline and follow-up and/or mean change values of the relevant outcome.

Statistical analysis

Data handling

Authors will use algebraic manipulation and imputation methods in order to reduce heterogeneity in data expression and include a higher number of studies. Indeed, not only articles reporting information as "mean change" will be evaluated for the analysis, but also papers with data expressed as "baseline value" and "follow-up value". In the latter case, changes will be calculated as the arithmetic difference between baseline and follow-up values.

Algebraic manipulation will be used for those papers reporting variability measures expressed as standard error or confidence interval: here, formulas proposed by the Cochrane Handbook will be adopted to re-calculate the standard deviation. Moreover, two imputation methods will be used in case of papers with incomplete variability measures.

The first method will be applied to papers without any variability measures, neither for mean change or for scores at baseline/follow-up: in this case, the higher SD will be imputed from another article with the same scale of measurement. The second method will be used for papers with complete data at baseline and follow-up: in this case, to calculate the SD of the changes, the correlation will be imputed from another article.

Effect size measure

Since outcome measures are expressed using different scales, standardized effect sizes (ES) will be computed for the purposes of the meta-analysis. The Glass' Delta index has been chosen as ES measure for this meta-analysis. Glass' Delta is appropriate in case of potential differences in variances among experimental and control groups. Glass' Delta uses the SD of the control group (or baseline comparison group) as the standardizer for the difference between the two treatment means. The reason behind this approach is that the SD of the control group is untainted by the effects of the treatment and will be therefore closer to the population standard deviation. The strength of this assumption is directly proportional to the size of the control group. The larger the control group, the more it is likely to resemble the population from which it was drawn. Glass' Delta reduces to Cohen's d or Hedge's g in case of equality of variances between experimental and control group.

Network model

A Bayesian multiple treatment network meta-analysis, with random effect and non-informative prior, will be performed. Between studies, standard deviation will be modeled using a uniform distribution on the 0-5 interval.

Random effects consistency model will be computed by Markov Chain Monte Carlo methods with Gibbs sampling based on simulation of 200000 iterations in 4 chains. The consistency assumption will be checked, wherever possible, following the method introduced by Dias. For each iteration, treatments will be ranked following their effect relatively to an arbitrary baseline. A frequency table will be constructed from these rankings and normalized by the number of iterations giving the rank probabilities. The convergence will be assessed using Gelman and Rubin diagnostic, the ratio between and within chain variation.

Probability values will be summarized and reported as surface under the cumulative ranking curve (SUCRA) and a rankogram plot, in order to provide a hierarchy of the treatments considering both the location and the variance of all relative treatment effects. The SUCRA is a measure providing a summary statistic for the cumulative ranking: the AUC (Area Under Curve) is the estimation of the surface underneath the cumulative ranking line. SUCRA would be 0 when a treatment is certain to be the worst and 1 when a treatment is certain to be the best. The direct and indirect effect for each treatment will be reported in a league table.

The analysis will be implemented using the R-evolution (3.3.1) software (with gemtc and R2OpenBugs packages) interfacing to Openbugs (3.2.3) for computing a Markov Chain Monte Carlo simulation.
Sensitivity analyses

Sensitivity analyses will be conducted in order to evaluate possible result variations as a consequence of placebo route of intra-articular placebo, which seems to be associated with small clinical benefit.35 On the other side, blinding process is a key-element in study design, especially for a trial investigating a patient-reported outcome: a lack of blinding in treatment administration can influence the perceived effects of the same treatment.

Following this consideration, researchers will conduct three different clinical sensitivity analyses:

- the first analysis will consider only studies with oral or intramuscular placebo, as well as no treatment;
- the second analysis will include only trials with intra-articular placebo;
- the third analysis will exclude all non-blinded studies.

Since the handling of missing data and the use of imputation methods could influence results, further statistical sensitivity analyses will be carried out in order to improve reliability of the estimations. For this reason, researchers will apply two statistical sensitivity analyses that gradually exclude imputation methods:

- the first analysis will exclude papers using the maximum SD imputation;
- the second analysis will exclude papers using the maximum SD imputation as well as papers applying the correlation imputation.

Further details of statistical analysis

Data processing will include all the algebraic manipulations and the imputation methods applied during the analyses in order to include treatments with incomplete data. Previous literature on meta-analysis suggests two main approaches to manage missing data: researchers can choose to select only articles with complete information or, otherwise, to adopt a specific imputation method. Although the first approach could seem more rigorous and accurate, it is not free of potential bias. The decision does not take in consideration relevant papers and can cause a reduction in the sample size and in a remarkable reduction in test power. Moreover, when a researcher chooses this type of approach, implicitly chooses to consider all data as MCAR (Missing Completely at Random): actually, probability of this event is really low, and when MCAR assumptions are violated the analysis will produce bias estimates. When we are considering studies with pharmacological treatments, this aspect has high importance, and should be taken into consideration in order to give exhaustive and correct clinical information. A meta-analysis on treatments comparison that includes just complete data, automatically excludes all papers in which incompleteness of data are probably due to non-significant results. For this reason, we choose to apply imputation methods in the main analysis and in the clinical sensitivity analyses. In the statistical sensitivity analyses, imputation methods will be progressively excluded.

Algebraic manipulation of data

The algebraic manipulations of the data will be applied in all the analyses, since data handling is widely accepted by the scientific literature and suggested by the Cochrane Collaboration. These methods will include:

- the transformation of Standard Error (SE) in Standard Deviation (SD): $SD = \sqrt{N} \times SE$
- the transformation of Confidence Interval (CI) in Standard Deviation: $SD = \sqrt{N} \times (CIU - CIL)/3.92$

Imputation methods

Two main imputation methods will be applied, the maximum SD imputation and the imputation of correlation, in three different cases:

I. in papers reporting only mean values at baseline and follow-up without any variability measures (SD or SE or CI),
II. in papers reporting only mean change value (Delta) without variability measures,
III. in papers reporting mean values at baseline and follow-up and variability measures.

In the first and second cases, the maximum SD imputation with or without imputation of correlation will be applied, while in the third case, we will apply just the imputation of correlation. If delta value are available, we will calculate it as the arithmetic difference between the mean value at follow-up and the mean value at baseline. In case the number of subjects related to delta is not available, we will use the number of subjects reported at follow-up.

Maximum SD imputation – consistently with this approach, we will impute the higher SD to each mean or delta value without variability measures (I and II cases). We will use the SD extracted from another study, following the below reported rules:

1. both standard deviations must have the same scale and range of measurement;
2. both standard deviations must be relevant to the same name of treatment (or placebo);
3. if more than one standard deviation is available, the higher SD will be chosen;
4. If SD is not available according to condition 1 and 2 and the treatment without SD is an NSAID, we will impute the maximum SD of another NSAID;

5. If condition n.4 is not satisfied, then we will impute the maximum SD of any other treatment or placebo.

**Imputation of correlation** – We will apply this approach to those trials with mean values at baseline and follow up and with variability measures (III case), as well as to those trials included in first case (I case) after the imputation of maximum SD. We will use the correlation extracted from another paper to calculate the SD for the delta value. The extracted correlation must have the same scale and range of the paper with missing data. We will extract correlations from studies with complete data only, which have both delta and mean values, and variability measures:

\[ SD_d = \sqrt{SD_b^2 + SD_f^2} - 2 \cdot corr(SD_b \cdot SD_f) \]

**Network meta-analysis (NMA) model description**

A Bayesian multiple treatment network meta-analysis, with random effect and uninformative prior, will be performed. The method is based on multiple comparison of treatment effect across different trials. Assuming to consider M trial comparing treatments 1 and 2, the treatments effects may be synthetized in a Fixed effect (FE) or in a Random effect (RE) meta-analysis model. A FE model assumes that each study provides the same treatment effect \( d_{12} \) and the estimate is affected only by sampling error. A RE meta-analysis assumes that the study specific treatment effect is not fixed but is a normal random variable with mean \( d_{12} \) and variance \( \sigma_{12}^2 \).

The trial specific estimate is exchangeable because it is obtained from a common random variable for all the trials comparing treatments 1 and 2. A fixed effect model is a special case in which the variability component, \( \sigma_{12}^2 \), is equal to 0.

The treatment effect may be expressed on different scales, continuous, binary, etc. The structure of the model is the same, but the link function and likelihood varies according to outcomes characteristics. The model is defined in a GLM theory leading to apply a link function \( g(\gamma) \) useful to mimic the support of dependent variable on \( \pm \infty \).

The treatment effect in study \( i \) for drug \( k \) may be expressed as:

\[ g(\gamma) = \theta_{ik} = \mu_i + \delta_{i,bk} \]

where \( g(\gamma) \) is the link function, \( \theta_{ik} \) is the study specific treatment effect, \( \mu_i \) is the baseline trial treatment effect associated to control treatment or treatment 1. \( \delta_{i,bk} \) is the k treatment effect compared to control treatment \( b=1 \), then in a RE model.

\[ \delta_{i,1k} \sim N(d_{1k}, \sigma_{1k}^2) \]

The model structure for a pairwise meta-analysis may be easily extended to a multiple treatment comparison (MTC). Assuming that also the treatment 3 has been included, then the study specific treatment effect of treatment 3 compared to control drug is assumed to be obtained (in a RE model) as a realization of a Normal random variable defined as \( \delta_{i,13} \sim N(d_{13}, \sigma_{13}^2) \). Then the estimates of the effects of treatment 2 vs 3 is easily obtained by difference as, considering that the study specific comparison in treatment effects are exchangeable:

\[ \delta_{i,23} = \delta_{i,13} - \delta_{i,12} \]

It is possible to take into account that the estimates provided in comparison between treatment 1 and 3, and treatment 1 and 2 are correlated then it is possible to obtain that:

\[ \sigma_{23}^2 = \sigma_{12}^2 + \sigma_{13}^2 - 2 \rho_{23}^{(1)} \sigma_{12} \sigma_{13} \]

For simplicity is assumed the homogeneity of variance then

\[ \sigma_{12}^2 = \sigma_{13}^2 = \sigma_{23}^2 = \sigma^2 \]

The \( \sigma^2 \) component represents the heterogeneity, specifically the between trial variability component.
Generalizing for \( l > 2 \) treatments to compare, the generalized consistency equation, as showed above, is based on the assumption of an overall exchangeability of treatments effect across trials.

\[
\begin{align*}
  d_{23} &= d_{13} - d_{12} \\
  d_{24} &= d_{14} - d_{12} \\
  \vdots \\
  d_{(s-1)} &= d_{1,s} - d_{1,s-1}
\end{align*}
\]

The NMA model leads to include multi arm trials. In this case, a single multi arm trial provides estimates that are correlated. Then assuming homogeneity of variances, as described above, estimates of correlated \( \delta_i \) are provided in a multinormal distribution:

\[
\delta_i = \begin{pmatrix} \delta_{i,12} \\ \vdots \\ \delta_{i,1a_i} \end{pmatrix} \sim N_{a_i-1} \left( \begin{pmatrix} d_{i1,t2} \\ \vdots \\ d_{i1,ta_i} \end{pmatrix}, \begin{pmatrix} \sigma^2 & \sigma^2/2 & \ldots & \sigma^2/2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^2/2 & \sigma^2/2 & \ldots & \sigma^2 \end{pmatrix} \right)
\]

**Model estimation**

Between study standard deviation will be modelled using a uniform distribution on the 0-5 interval. Random effects consistency model will be computed by Markov Chain Monte Carlo methods with Gibbs sampling based on simulation of 200000 iterations in 4 chain. The consistency assumption will be checked, wherever possible, following the method introduced by Dias.\(^34\) For each iteration, the treatments will be ranked following their effect relative to an arbitrary baseline. A frequency table will be constructed from these rankings and normalized by the number of iterations giving the rank probabilities. The convergence will be assessed using Gelman and Rubin diagnostic, the ratio between and within chain variation.\(^30\)

Probability values will be summarized and reported as surface under the cumulative ranking curve (SUCRA) and a rankogram plot, in order to provide a hierarchy of the treatments considering both the location and the variance of all relative treatment effects. The SUCRA is a measure providing a summary statistic for the cumulative ranking: the AUC (Area Under Curve) is the estimation of the surface underneath the cumulative ranking line.\(^32,33\) SUCRA would be 0 when a treatment is certain to be the worst and 1 when a treatment is certain to be the best. The direct and indirect effects for each treatment will be reported in a league table.

The analysis will be implemented using the R-evolution (3.3.1) software (with gemtc and R2OpenBugs packages) interfacing to Openbugs (3.2.3) for computing a Markov Chain Monte Carlo simulation.


Methodology

Information sources and search procedure

The search procedure on the electronic databases was extended to retrieve full articles until June 30th 2018, with a
National Library of Medicine weekly alert until August 31st 2018.

Study quality was assessed by the Jadad score but the data have not been presented in the publication, where only the
Cochrane Collaboration’s tool for assessing risk of bias has been presented.

Statistical analysis

Effect size measure

An additional analysis of the effect size on the primary outcome (pain) has been performed normalizing WOMAC
scores to a 0-100 scale. This provides a measure comparable to a single question VAS pain assessment (0-100 mm).
Results of the primary analysis were therefore provided also on the normalized 0-100 scale.

Further details of statistical analysis

In order to comply with the peer reviewers’ requests for publication, an Additional Statistical Details section has been
provided in the publication material to better explain the statistical methodology employed and comply with the request
of different post-hoc sensitivity analyses.
All the post-hoc sensitivity analyses as well as those described in the study Protocol have been implemented using the
R-evolution\textsuperscript{1,2} (3.3.1) software (with gemtc\textsuperscript{3} package) interfacing to OpenBUGS\textsuperscript{4} (3.2.3).

URL \url{http://www.rstudio.com/}
0.8. gemtc: Network Meta-Analysis Using Bayesian Methods R package.
The Cochrane risk of bias tool is based on the following domains in the original article:

- **Random sequence generation** Description of the method used to generate the allocation sequence in sufficient detail to allow an assessment of whether it should produce comparable groups.

- **Allocation concealment** Description of the method used to conceal the allocation sequence in sufficient detail to determine whether intervention allocations could have been foreseen in advance of, or during, enrolment.

- **Blinding of participants, personnel and outcome assessors** Description of all measures used, if any, to blind study participants and personnel from knowledge of which intervention a participant received. Provision of any information relating to whether the intended blinding was effective.

- **Incomplete outcome data** Description of the completeness of outcome data for each main outcome, including attrition and exclusions from the analysis. Statement whether attrition and exclusions were reported, the numbers in each intervention group (compared with total randomized participants), reasons for attrition/exclusions where reported, and any re-inclusions in analyses performed by the authors.

- **Selective outcome reporting** Statement of how the possibility of selective outcome reporting was examined by the authors.

- **Other potential threats to validity** Statement of any important concerns about bias not addressed in the other domains in the tool.

Within studies, each of the above mentioned domain can be judged according to the following Risk of bias Interpretation:

- **Low risk of bias** Plausible bias unlikely to seriously alter the results.

- **Unclear risk of bias** Plausible bias that raises some doubt about the results.

- **High risk of bias** Plausible bias that seriously weakens confidence in the results.

In addition, for each study the overall risk of bias was assessed as following:

- **Low**: if risk of bias was low in all first four domains (key domains)

- **Unclear**: if risk of bias was unclear in 1 or 2 of the four key domains

- **High**: if risk of bias was high in ≥1 or unclear in ≥3 of the key domains