Author's response to reviews

Title: Comparison of intra-articular injections of Hyaluronic Acid and Corticosteroid in the treatment of Osteoarthritis of the hip in comparison with intra-articular injections of Bupivacaine. Design of a prospective, randomized, controlled study with blinding of the patients and outcome assessors.

Authors:

Sascha Colen (sascolen@hotmail.com)
Michel P. J. van den Bekerom (bekerom@gmail.com)
Johan Bellemans (secretariaat.bellemans@uz.kuleuven.ac.be)
Michiel Mulier (michiel.mulier@uz.kuleuven.ac.be)

Version: 5 Date: 11 September 2010

Author's response to reviews: see over
Dear Editor and Reviewers,

Thanks for the high quality and well documented comments. We have changed and updated the manuscript according these comments and advices. All authors have read and approved the manuscript. For our answers (in bold and italic) and changes of the manuscript, see below.

Best regards,

The authors

Please note that a competing interests section must be included in your manuscript before we consider your revised manuscript further.

In 2010 the authors received funding from Ostenil® (TRB Chemedica, Haar, Germany) to pay article-processing charge for publishing their article in BMC musculoskeletal disorders. This funding company has absolutely no role in study design; in the collection, analysis, and interpretation of data; in the writing of the manuscript; and in the decision to submit the manuscript for publication. Neither they nor a member of their immediate families received payments or other benefits or a commitment or agreement to provide such benefits from a commercial entity. No commercial entity paid or directed, or agreed to pay or direct, any benefits to any research fund, foundation, division, centre, clinical practice, or other charitable or non-profit organization with which the authors, or a member of their immediate families, are affiliated or associated.

We recommend that you copyedit the paper to improve the style of written English. The style of written English is improved by a native English speaker.

Reviewer’s report

Title: Comparison of Hyaluronic Acid and Corticosteroid intra-articular injections for the treatment of Osteoarthritis of the hip controlled with intra-articular injections with Bupivacaine. Design of a double blinded prospective randomized controlled study.

Version: 3 Date: 17 June 2010

Reviewer: Damien Jolley

Reviewer’s report:
The manuscript presents a proposed protocol for a RCT planned to compare Hyaluronic acid (HA) to corticosteroids (CS) with a control group receiving Bupivacaine (B) by a similar intra-articular route for treatment of osteoarthritis (OA) of the hip.

First, the authors are to be commended for the decision to publish their protocol
before the study is conducted. As they remark, this will ensure the rigour of the RCT, by having to meet published methods. The Cochrane Collaboration insists on publication of peer-reviewed protocols for all of their systematic reviews. The manuscript describes a reasonably rigorous RCT design, but there are a number of omissions, and I have a number of potential improvements to suggest. The editors of BMC Musculoskeletal will address the language issues in the manuscript, but I have to state that these are numerous, and certainly need attention. I will not list them in this review.

The four principal issues of bias for Cochrane reviews of RCTs are: (1) generation of the randomisation sequence; (2) concealment of allocation; (3) effective blinding; and (4) risk of incomplete outcome data reporting (http://www.cochrane-handbook.org). We address these issues in this review.

1) Randomisation will be performed using www.randomizer.org. This is suitable for sequence generation, but it seems a set of printed lists will be generated and then downloaded from the site. This has implications for concealment of allocation (see below).

2) If lists are printed and available for view to the allocating nurse, then concealment is not possible. An alternative mechanism for the randomisation process could solve this – for example, using opaque envelopes, or real-time randomisation at infiltration.

3) The authors state that the RCT will be double-blinded, because neither the patient nor the “post-operative” data collectors will be aware of the allocation. This is not the usual meaning of “double-blind”, however. See, for example, Shulz & Grimes:

In a double-blind trial, participants, investigators, and assessors usually all remain unaware of the intervention assignments throughout the trial. In view of the fact that three groups are kept ignorant, the terminology double blind is sometimes misleading. In medical research, however, an investigator frequently also assesses, so in this instance the terminology accurately refers to two categories.

THE LANCET • Vol 359 • February 23, 2002, p697

In this protocol, it appears that the investigators/surgeons will be aware of the assignment of each patient. Using Shulz & Grimes definition, this is not double-blinded. There are good reasons to blind the investigators/surgeons – knowledge of the allocated treatment may consciously or subconsciously bias the carer to better care.

5) The recognised standard for RCT data analysis is intention-to-treat (ITT). In a nutshell, this can be stated: “Patients are analysed in the groups to which they are randomised.” It is extremely concerning, therefore, to see this protocol’s third
Exclusion Criterion: “Non-compliance to the study procedures and or non-completion of the study according to investigator’s judgement.” This criterion makes ITT analysis impossible. Patients must be included and analysed whether they follow the treatment protocol or not. Further, the exclusion criterion shows the importance of having investigators/carers blind to the allocated treatment. If the investigator can unilaterally choose to exclude any patient, there is an exceedingly high risk of selection bias.

From the above list, it is clear that this protocol needs some revision to avoid a study with high risk of bias. There are some other issues concerning the proposed methods and sample size.

I shall begin with the hypotheses, which are stated in terms of binary results, for example “There are differences in functional outcome (HHS & VAS) between the treatment groups (HA & CS) and the control group (B) in favour of the treatment groups, for the treatment of OA of the hip at 6 months follow-up”. A protocol needs to specify the magnitude of the minimum clinically important difference. If you choose to state the alternative hypothesis, rather than the null hypothesis, then it is your responsibility to nominate what you mean by “differences” (and just “significant differences” is invalid as well). This failure in the hypotheses statement has ramifications later when considering sample size. When sample size is considered (under “Description methodology”), the authors choose to use a dimensionless standardised effect size of delta=0.4 to obtain their result. But this number has no clinical meaning at all. It is defined as the difference in observed means, divided by the standard deviation of the population (as in psychometric research) but does not communicate whether the patient will feel better or perform better. Only the ABSOLUTE difference in means can achieve this – for example, we declare that a change in mean VAS of 10mm or greater will be considered to be clinically important. (Although some studies report “SMD” as defined in the Cochrane Handbook, this is recommended only when comparing studies which have different outcome measures. Within the one study, with identical outcome measures, the difference in means is the appropriate effect size for continuous measures.) The protocol’s sample size calculations are correct, assuming the stringent alpha=0.01, power=0.8 and delta=0.4, but once again ignoring the need for ITT analysis, the protocol expects to obtain no follow-up data from about 10% of patients.

The statistical analysis plan is vague. A “linear regression model for repeated measures” is probably best replaced by a more flexible method such as that used by Qvistgaard et al (ref 15). The plan fails to mention the reporting of confidence intervals at all. In fact, confidence interval analysis is preferred throughout, making Bonferroni adjustments and interim analysis P-values unnecessary. To summarise, this protocol describes an RCT which is at risk of bias and has unclear power calculations. The document’s English expression needs attention in several places. The authors are to be commended for subjecting their protocol
Major compulsory revisions:
1) Ensure ITT analysis
   We fully agree with this reviewer that one of our exclusion criteria has to be deleted.
   This criterion (Non-compliance to the study procedures and or non-completion of the study according to investigator’s judgment) is deleted from our protocol now.

2) Revise sample size to reflect absolute effect sizes
   We fully agree with the reviewer that a specification of the difference expressed on the scale of the used measurement is clinically more relevant than using the dimensionless standardized effect size. To perform the power analysis, two sources of information are then required: the (unstandardised) effect size which is expected and deemed clinically relevant and an estimate of the standard deviation. It is not trivial to provide respectively a precise definition and estimate. However, we followed the recommendation of the reviewer and (after a check in the literature and a discussion amongst the authors) we defined a difference of 10mm in VAS and a difference of 10 points in HSS to be clinically relevant. We obtained an estimate of the SD of the HSS from Frihagen et al. (2008), being equal to 16 (pooled estimate of HSS after 4 months, Table 3) and an estimate of the SD for VAS equal to 20 (from Frihagen et al, 2008 and Qvistgaard, 2006). To calculate the SD of the change in HSS and VAS (needed for the power analysis, since the evaluation of the primary outcomes is based on a comparison of the changes between groups), an estimate of the correlation between the time points is needed. Note that a correlation of 0.5 will result in a SD of the difference which equals the SD of the baseline measurement. Assuming a moderate correlation of 0.50, a sample size calculation is performed to have in both groups at least 80% power for both outcomes to detect a difference with the placebo group. In total 285 patients are needed, i.e., 95 subjects in each group. With this sample size, 80% power is achieved for VAS and 96.2% for the HHS.

   We have redone the sample size calculation and rewritten the part referring to the sample size calculation as follows:

   “A sample size calculation is performed to have at least 80% power to detect in both groups compared with placebo, a clinically meaningful difference in change (baseline-6months) of 10mm in VAS and 10 points in HSS. The SD of HSS and VAS is assumed to be equal to 16 and 20 respectively. Estimates are obtained from Frihagen et al. (2008) and Qvistgaard (2006). Assuming a moderate correlation of 0.5 between the measurement at baseline and after 6 months, in total 285 patients are needed, i.e., 95 subjects in each group. With this sample size, 80% power is achieved for VAS and 96.2% for the HHS”.

   We think that in our hospital 10 % lost to follow-up is realistic because people can be placed on the waiting list for a total hip arthroplasty during follow up and the
operation will be performed six months after the initial operation. It is expected that most patients do not change hospital or treating physician in our district.

3) Commit to reporting confidence intervals, rather than P-values

*We agree that confidence intervals should always be given.* We failed to mention this in the methodology. We have added the sentence: “Confidence intervals (CI) of the change as well as the difference (at each time point) between groups will be constructed.”

A conclusion (rejection of the null hypothesis or not) can indeed easily be drawn based on the CI, as long as the test has a single degree of freedom. But also for the construction of a CI adjustments, similar as those in the proposed methodology, are still needed to handle the issues of multiple testing and multiple looks at the data. For example, for the construction of a CI based on the standard normal distribution, the z-values 4.495, 3.178 and 2.595 will be used at the three analysis moments. As explained in the methodology, this corresponds to declaring p-values significant at .000007, .00148 and 0.0095 level.

*We realize that confidence intervals and p-values are strongly linked, making p-values (partially) redundant. Also since not reporting p-values is not (yet) common practice, we deem it most appropriate to report them both.*

Minor Essential Revisions
1) Address allocation concealment

*We will use computer randomization rather than making a printed list with the numbers because this will reduce the risk of bias. This will be performed real time at the time of the infiltration.*

2) Revise blinding requirements

*To our knowledge there is no uniformity of these definitions, maybe the term triple blinded would be necessary??). Our new title will be: Comparison of Hyaluronic Acid and Corticosteroid intra-articular injections for the treatment of Osteoarthritis of the hip controlled with intra-articular injections with Bupivacaine. Design of a prospective randomized controlled study with blinding of the patients and outcome assessors.*

Discretionary Revisions
1) Consider rewording of hypotheses

*There is a difference in Visual Analog Score (VAS) between the treatment groups (HA and CS infiltration) and the control group (Bupivacaine) of 10 mm in favour of the treatment groups, for the treatment of OA of the hip at 6 months follow-up.*

2) Consider improved analysis methods 3) Revise stringent alpha (95% confidence intervals only are necessary).

*The protocol’s sample size calculations are correct, assuming the stringent alpha=0.01, power=0.8 and delta=0.4, but once again ignoring the need for ITT analysis, the protocol expects to obtain no follow-up data from about 10% of patients. The description of the analysis has been adapted, explicitly stating that*
an intent-to-treat analysis will be performed, i.e. all subjects receiving an infiltration are analysed in the primary analysis, whether or not they have completed the total follow-up. The number of included subjects will be increased with 10%, not to replace the patients with incomplete follow-up, but to compensate for the loss of power. This loss occurs since some patients are expected to have one or more missing observations. Note that the last-observation-carried forward (LOCF) technique applied by Qvistgaard et al. inappropriately ignores this loss of power by performing a (disputable) single imputation. If an imputation technique is used, it should be a multiple imputation approach. Instead, we used a direct likelihood method (see infra), resulting in standard errors which will reflect the increased uncertainty (loss of precision) due to missingness. Whether a multiple imputation approach or a direct likelihood approach is used, the planned sample size should compensate for expected presence of missingness.

Level of interest: An article of importance in its field
Quality of written English: Needs some language corrections before being published
Statistical review: Yes, and I have assessed the statistics in my report.
Declaration of competing interests: I declare that I have no competing interests.

Reviewer's report

Title: Comparison of Hyaluronic Acid and Corticosteroid intra-articular injections for the treatment of Osteoarthritis of the hip controlled with intra-articular injections with Bupivacaine. Design of a double blinded prospective randomized controlled study.

Version: 3 Date: 24 May 2010

Reviewer: Burkhard Leeb

Reviewer's report:
Major compulsory revisions:

Given the text the reviewer assumes that the primary endpoint of the study should be an improvement of the HHS in the HA and CS treated patients in comparison to the controls (Bupivacain) at month 6. To the reviewer’s knowledge there are insufficient data supporting a six month efficacy of CS injections into the hip. However, no difference between HA and CS injection is obviously expected by the authors.

The H0 hypothesis is that there is a difference between the treatment groups and placebo. There could be a difference between HA and CS (as you already stated this is also expected) but this is not stated in the H0 and the sample size is subsequently not calculated based on this last hypothesis.
As far as the reviewer understood the protocol, one injection is intended per patient included. As the efficacy of a single HA injection into a joint is also not yet proven (even for knee joints), it remains doubtful whether it will be possible to meet this primary endpoint of six months.


First, the authors should indicate a clear primary endpoint and a distinct primary outcome measure (e.g. the HHS or the VAS) of the trial.

The sample size was calculated on this primary endpoint / H0 hypothesis: There is a difference in Visual Analog Score (VAS) between the treatment groups (HA and CS infiltration) and the control group (Bupivacaine) of 10 mm in favour of the treatment groups, for the treatment of OA of the hip at 6 months follow-up.

Second, the reviewer would propose to reconsider, whether six months after the application of the study drugs would constitute an appropriate primary endpoint. In our opinion is this minimal duration that is required to justify the potential complications. A treatment effect of 3 weeks is clinically not relevant and should not justify the risk of nerve/vessel damage, bleeding or infection.

One of the inclusion criteria is chronic pain for the last three months. This statement in fact seems to be rather vague. A decisive value on the VAS pain for patients to be included is seriously proposed; e.g. 30 mm on the VAS at screening, and no improvement at the day of the injection. In addition, a threshold value on the HHS or the HOOS at the screening visit for patients to be included should be given.

We agree that a threshold in VAS is relevant. We used a threshold of 30 mm on the day of inclusion in our study. Using a threshold HHS and HOOS as inclusion criteria is not realistic in our clinical practice.

To assume an ES of 0.4 for HA and CS treatment comes near the ES Zhang et al. found for PBO intra-articular injections in their meta-analysis of PBO responses in OA clinical trials. Therefore a reconsideration of the ES is strongly recommended. In addition, an anticipated drop-out rate of 10% can be regarded very optimistic in a six months hip OA trial. Therefore, the reviewer would recommend to reappraise the drop-out rate, and to increase the number of subjects to be included.

As discussed above, see second major compulsory revision from the other reviewer.

The writing can be regarded acceptable.

Level of interest: An article of importance in its field
Quality of written English: Acceptable
Statistical review: Yes, but I do not feel adequately qualified to assess the
statistics.

**Declaration of competing interests:**
Dr. Leeb has received consultancy fees from TRB Chemedica, and IBSA. Dr. Leeb was involved in clinical trials sponsored by IBSA. Dr. Leeb has received speaker’s honoraria from TRB Chemedica, IBSA, Lacer-Spain and Biosaude-Portugal.