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: 4 Date: 5 September 2010

Author’s response to reviews:

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 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.

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 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.


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 HHS and HOOS as inclusion criteria is not realistic in our clinical practice. The patients will be included in the outpatient clinic so time is sparse and performing this study will be time consuming.

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.

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 timepoints 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 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.

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 timepoint) 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.

“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).”

We agree that the description of the statistical plan should contain more details about the statistical technique. We have added the following paragraph describing the proposed model and including references:

“More specifically, a direct likelihood approach is adopted using an unstructured covariance matrix for the repeated measures (Molenberghs and Kenward, 2007,
Section 14.4). The approach has the advantage that less stringent assumptions are made, not only with respect to the covariance structure, but also to the mechanism underlying the missing observations. The approach neither assumes all variances to be equal, nor all covariances (as it done in classical repeated-measures ANOVA or in a mixed model using only a random subject effect). Further, the approach allows that the missingness depends on the observed values (the so-called missing at random assumption (MAR)), whereas an analysis restricted to patients with complete information assumes the missingness to be completely at random (MCAR) (G. Molenberghs and M.G. Kenward, 2007).


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.

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.

Consider rewording of hypotheses
Clear primary endpoint and a distinct primary outcome measure

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.

Consider improved analysis methods / 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 analyzed 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.