Author's response to reviews

Title: Efficacy of zoledronic acid for chronic low back pain associated with Modic changes in magnetic resonance imaging

Authors:

Katri Koivisto (ka.koivisto@elisanet.fi)
Eero Kyllönen (eero.kyllonen@oulu.fi)
Marianne Haapea (marianne.haapea@oulu.fi)
Jaakko Niinimäki (jaakko.niinimaki@oulu.fi)
Kaj Sundqvist (kaj.sundqvist@ppshp.fi)
Timo Pehkonen (timo.pehkonen@ouka.fi)
Seppo Seitsalo (seppo.seitsalo@orton.fi)
Osmo Tervonen (osmo.tervonen@oulu.fi)
Jaro Karppinen (jaro.karppinen@ttl.fi)

Version: 2 Date: 24 January 2014

Author's response to reviews: see over
REPLY TO REVIEWERS

Re: MS: 1359258775112375 - Efficacy of zoledronic acid for chronic low back pain associated with Modic changes in magnetic resonance imaging

Dear Editor,

The authors would like to thank you and the Reviewers for your time and effort devoted to the review of our aforementioned manuscript. The comments were insightful and greatly appreciated. As such, the authors would like to take this opportunity to address each concern the Reviewers noted in their review of our submission. In addition, where appropriate, we have revised our manuscript accordingly.

The additions to the manuscript are shown in red. Unfortunately, deleted text is not shown in order to improve the readability of the manuscript.

We have changed the adjustments of the analyses according to the comments of external experts. See response to comment 4.

We have also clarified the role of the funder in the Financial support section on page 1.

We believe that the Reviewers’ comments have helped improve the quality of our manuscript. We hope that you and the Reviewers will now find our work suitable for publication in the BMC Musculoskeletal Disorders.

Reviewer 1: Professor Charlotte Leboeuf-Yde

1. Is the question posed by the authors well defined?  
   Yes
2. Are the methods appropriate and well described?  
   Yes, overall. However, some comments:

   Comment 1: No need to say in Methods under Treatment intervention, 2nd sentence that the study was “blinded”. It is not completely explained at this point. As a reader, I immediately started to wonder how it was blinded. An excellent explanation of that comes later, so I suggest remove the word “blinded” the first come it appears and save it for later to avoid early confusion.

   Response: We have removed “blinded” as suggested.
Comment 2: Please describe what your definition of clinically acceptable improvement is, and the rationale for this.

Response: We thank the Reviewer for the comment. We have used 20% relative improvement as minimal clinically important improvement as suggested by Tubach and colleagues (Tubach et al. Arthritis Care Res 2012; 64:1699-707). We have added this to the Methods section (p. 7). Additionally, we use “patient acceptable symptom state (PASS)” as an additional measure of improvement as suggested by Tubach et al. (the proportion of patients reaching 4 or less on 10-cm VAS scale, which is regarded as acceptable status). These results are shown in the results (p. 8, end of the first paragraph of ‘Treatment differences’), but only for the primary outcome.

Comment 3: It is not acceptable to look at estimates and describe all such estimates that go in the “right” direction as “tendencies” or “trends”. You need a pre hoc “guideline” for this. To pronounce a “trend”, you need to do a test of trend!

Response: We agree with the reviewer that trends/tendencies are inappropriate terms and have removed them.

Comment 4: You are performing at least 40 statistical tests. With p at 0.05, you will probably obtain at least two “falsely” significant values, so either you have to explain this very clearly in your discussion section, or you have to adjust your p-value with the Bonferroni method (0.05 divided by 40) or do something else to accommodate to this fact.

Response: We agree with the Reviewer that multiple testing may have occurred but only in secondary outcomes. The primary outcome had been defined a priori, and does not need correction. We have added on p. 12 (3rd paragraph) in the Discussion section: “However, due to multiple testing, the significance levels of the secondary outcomes must be interpreted with caution.”

Furthermore, after reconsidering the need for adjustments, after discussing with experts in methodology (see acknowledgment for names), we omitted adjustments for age and gender because any differences in baseline characteristics are the result of chance rather than bias (http://www.consort-statement.org/consort-statement/13-19---results/item15_baseline-data/). In fact, adjustment with them did not change point estimates – only widened confidence intervals as stated on p. 7 in the Methods section. Now we report only results after adjustment with the baseline score of the respective variable, in addition to the unadjusted results. This has simplified Tables 2 and 3.
Comment 5: In the statistical analysis text, explain what you do with normally and not normally distributed data.

Response: We thank the Reviewer for the observation. Within background variables the duration of LBP, leg pain, and sick leave were skewed, so we described them using median values with interquartile range instead of mean values with standard deviation. This has been clarified on p. 7 in the Methods section.

The treatment differences were analysed using parametric methods in order to be able to adjust for the baseline score. The distributions were mainly symmetrical. The results of unadjusted analyses would not have been different if non-parametric Mann-Whitney’s test had been used.

3. Are the data sound?
Yes, as far as it is possible to judge that without having access to raw data. All looks reasonable though.

4. Does the manuscript adhere to the relevant standards for reporting and data deposition?

Comment 6: Yes, but the Result section is messy and your results are embarrassingly over interpreted. I have some comments.

Response: We have revised the manuscript and hope that it is less over interpreted now.

Comment 7: Report first on your primary outcome variable. Your research question was if there was a difference in outcome between treatment A and B. Then please report clearly on this, using a clinically relevant cut point as your point of demarcation. Show some courage; It is OK to report that it did not work the way you thought it would! Thereafter, explain in words, if there were any clinically and statistically significant findings without trying to “fool” the reader with trends and tendencies. Either it is significant or it is not. If not, then just say that the other variables did not achieve significance. We do not want to read about non-significant information. You stated your level of significance in the Methods section, so please stick to it. Also, you have provided exact estimates in tables, so there is no need to reiterate these (uninteresting) findings in the text. In fact, your only “real” results are the NSAIDs difference of 20% vs. 60%.

Response: We thank the Reviewer for the comment. We have revised the paper and hope that it is acceptable now.

Comment 8: You will have to choose if you want to report your results as tables or as graphs, you cannot use both methods for the same data.
Response: We thank the Reviewer for the suggestion. We have removed both figures as suggested.

Comment 9: Your Figs 2 and 3 (was there a Fig 1?) relate to data collected three times. Do not combine these points with lines. These lines falsely make the reader believe that you have continuous data, which you have not. You have three cross-sectional data collections, so to speak. A bar graph is what you need here or you could indicate the mean values with confidence intervals around (if it is confidence intervals you are showing; you forgot to tell us that).

Response: We thank the Reviewer for the observation. We have now removed the figures as suggested in comment 8.

Comment 10: Suddenly in the Result section you tell us that a 20% improvement in intensity of LBP tended to favour the ZA treatment etc. Why twenty percent? Please explain (here or previously in text).

Response: Please, see the response to comment 2.

5. Are the discussion and conclusions well balanced and adequately supported by the data?

Comment 11: Fairly well balanced, but I have some suggestions. There is lack of discussion on whether the results can be explained by the natural course, regression towards the mean and suchlike.

Response: We thank the Reviewer for a good suggestion. We have added some text (p. 10, second paragraph):

“The natural course of MC is not well known. Usually M1 lesions convert to M2 lesions with time [5], although small M1 lesions may also normalize [6]. According to the current view, the persistence of the M1 component correlates with persistence of symptoms [13,26]. We observed in another study population that symptoms persisted in almost one third of patients over a two-year follow-up, and that this persistence of symptoms was related to the persistence of the M1 component (Järvinen, unpublished observation). It is interesting to evaluate the course of symptoms in relation to changes in the M1 component on MRI in the current study population.”

Comment 12: You seem fairly hooked on this treatment, which I can understand because it is plausible. However, the small improved estimates that you see sometimes (not always, as you claim), could well have occurred because most patients figured out that
they were in the treatment group. Almost all of the real treatment patients had adverse
effects but only few in the other group. You need to discuss this possibility, and in
particular, you have to excuse yourselves for not having done a check of whether patients
guessed which group they belonged to!

**Response:** We thank the Reviewer for the comment. We have added to the discussion (p.
12, second paragraph):

“The patients, the study nurse, the medical team in charge of the patient, the
physician performing the assessments and infusion, and the statistician performing the
analyses were all blinded to the allocation. However, the high incidence of acute phase
reaction symptoms in the ZA group may have revealed the concealment to some patients.
Unfortunately, we did not evaluate the patients’ perception of the nature of the treatment
they had received.”

**Comment 13:** You did not do a power calculation. I accept that you did not. Now is the
time to make up for this weakness and to calculate how many study subjects you would
need in a “proper” study.

**Response:** We thank the Reviewer for the comment. We performed the power
calculations and noticed that power was very low in most of the analyses when calculated
using the actual estimates from the data. In the primary outcome, change in the intensity
of LBP, the power for unadjusted analysis at 1 month is circa 40% and at 1 year only
about 10%. The respective figures for adjusted analyses are 50% and 15%. In order to
achieve 80% power at 1 month in unadjusted analyses (with alpha = 0.05), we would
have needed 57 subjects in each treatment group, and at 1 month in adjusted analyses 38
subjects in each treatment group. We acknowledge that the study was underpowered.
However, despite this we had a significant treatment difference in the primary outcome at
1 month and in NSAID use at the one-year follow-up.

**Comment 14:** However, you will need to look at the 30% level, which is the usual
improvement level in studies on LBP. Would you then get a difference between group
estimates at all?

**Response:** Please, see the response to comment 2. We have used 20% relative
improvement as minimal clinically important improvement as suggested by Tubach and
colleagues. Additionally, we use “patient acceptable symptom state (PASS)” as an
additional measure of improvement as suggested by Tubach et al.

6. Are limitations of the work clearly stated?
Comment 15: Some are but the authors very clearly believe strongly in this method of treatment, their interpretation is really very optimistic.

Response: We have downgraded the ‘optimism’ as suggested. We hope the manuscript is acceptable now.

Comment 16: The major problem, in my view, is the possible lack of naïve study subjects. Because of the adverse effects, that most people in the “real” treatment group experienced, it would be easy for patients to assume they had been ”treated”, despite an excellent study design where everybody that could be blinded was blinded.

Response: Sorry, we do not understand the meaning of ‘totally naïve patients’. The enrolled patients had chronic LBP and MC in MRI. Adverse effects were reported by a considerable percentage of patients in the placebo group. We have no reason to believe that patients of the treatment group would have been de-blinded. We have, however, acknowledged this a limitation. See also the response to comment 12.

7. Do the authors clearly acknowledge any work upon which they are building, both published and unpublished?

Comment 17: I dislike reviewers who tell the authors to quote work of the reviewer. But the fact is that you did forget a recent RCT that I co-authored. It dealt with chronic LBP in people with Modic changes (Jensen KR et al) which basically studies the same aspect of Modic changes as you do; namely the quality of the bone. Rest vs. physical activity were the two test groups and the results were that there was no difference between the groups.

Response: We have added the reference Jensen et al. 2012 as suggested.

8. Do the title and abstract accurately convey what has been found?

Comment 18: The title explains what the study is about, the abstract deals with the results. The result section does not, in my opinion, give a entirely fair picture of the results.

MJCOMP: Clean up the result section in abstract by being hard an honest with your own results.

Response: We thank the Reviewer for the comment. We have brought forward the results with different emphasis. We hope that the revision has improved the manuscript.
9. Is the writing acceptable?
Yes, very good.

**Response:** The manuscript has been checked by a native English speaking professional.