Reviewer's report

Title: Caffeine is a Risk Factor for Osteopenia of Prematurity in Preterm Infants: A Cohort Study

Version: 1 Date: 25 May 2017

Reviewer: Neil Walker

Reviewer's report:

Overall, this comes over as an interesting and useful piece of research. I wasn't familiar with the subject area but nonetheless found the article accessible. My own background is in statistics and I do feel it would be necessary to revise parts of the methods and statistics sections for the purposes of clarity before proceeding to publication.

My suggestions are as follows:

- At different times the authors speak of "Generalized Linear Models" and "Generalized Mixed Models". The first of these I am familiar with, but the second I believe to be "Generalized Linear Mixed Model" (GLMM). This may appear a little pedantic, but the terminology is quite standard and to avoid confusion for the statistically aware reader I would recommend switching to this terminology (unless they did use a method called Generalized Mixed Models which I am not yet familiar with!).

- More clarity is required in the description of the GLM/GLMMs used. In particular... (i) a binary (0/1) response was analysed. How was it modelled? Would typically be with a binomial distribution and logit (possibly probit) link but this isn't stated. (ii) I understand that for any given child there are a number of OP measurements over 2 week periods. This calls for some kind of repeated measure analysis since you cannot assume that OP readings from the same individual are independent (indeed this is unlikely). Perhaps this is why the authors have gone down (what I assume is) the GLMM route. But authors need to be clear about how this was accounted for.

- Rather than fitting a multiple regression with all the significant main effects and then fitting a separate model with the interaction, I think the authors could make the presentation more concise by just concentrating on the interaction model (in addition to the univariate results) as this seems to cover all the main points arising from the main effects only model.

- Reporting of the results could be done more concisely, e.g. there's no need to list in the text all the variables which didn't show up as statistically significant. Also there's a degree of
repetition of results in the text that can be ascertained from the tables (though it's fair enough to do this for the most important results).

- There are some parts of the Results section which describe the evolution of the statistical methodology (e.g. top of page 15 which I believe is the first time an interaction is mentioned). Such descriptions of model choice would surely be better placed in the Methods section.

- Tables 4 & 5: model estimates are presented. But what scale are they on? I am assuming, given the GLM approach, that these are on the logit scale (though I might be wrong!). If so, this needs to be stated.

- The figures: are these graphs based on the raw data or on model estimates? I presume it's the latter, but I would recommend making this clear in the respective titles (if I'm correct in my assumption!). For all the figures I would use a vertical axis running from 0 to 1. Finally, these figures suggest that where the caffeine dosage is very low the probability of OP is negligible and when very high the probability is almost 1, i.e. a near certainty. I haven't done the calculation myself… just it seems a bit surprising to me that it would be such a dramatic effect. But then I'm new to the subject so maybe reasonable… My feeling is that it would be worth the authors double-checking this to be sure that the probabilities shown on these graphs have correctly mapped over from the model estimates.

- In the Discussion section, I would be careful not to rule out the presence of effects where a variable has not been found to be statistically significant as the sample size is not large enough to draw such strong conclusions.

- Overall I think the paper is well written although there are a few grammatical things which I assume would get picked up in the editorial process. One thing I would definitely change is the word "demographic" (bottom of page 7) - most/all of these variables are not demographic. Most/all of them are biometric I would say, but could just say "the data included…".

Whilst I have made a number of points here, I feel that most of them should be relatively straightforward to address and that the paper will make for a good read once this has been done.

**Are the methods appropriate and well described?**
If not, please specify what is required in your comments to the authors.

No

**Does the work include the necessary controls?**
If not, please specify which controls are required in your comments to the authors.

Yes
Are the conclusions drawn adequately supported by the data shown?
If not, please explain in your comments to the authors.

Yes

Are you able to assess any statistics in the manuscript or would you recommend an additional statistical review?
If an additional statistical review is recommended, please specify what aspects require further assessment in your comments to the editors.

I am able to assess the statistics

Quality of written English
Please indicate the quality of language in the manuscript:

Needs some language corrections before being published

Declaration of competing interests
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I declare that I have no competing interests

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors'
responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal