Reviewer's report

Title: General movements in preterm infants undergoing craniosacral therapy: a randomised controlled trial.

Version: 5 Date: 2 November 2015

Reviewer: Eric Jacobson

Reviewer's report:

Review of Raith et al. "General movements in preterm infants undergoing craniosacral therapy: a randomised controlled pilot trial"

Reviewer: Eric Jacobson

The ms. was again very much improved by the most recent round of editing and I believe almost ready for publication. However there is one final area of serious confusion that should be corrected. If this confusion is corrected then the article will be ready for publication.

Major required revision:

The way in which the linear mixed model and the p values derived from it are reported are very confusing and should be clarified. This confusion occurs in lines 197-98, 237-41, Table 2, and the Legend to Figure 2. Lines 21-23 in the Summary repeat these confusions.

Many of these confusions come about because of references to "time" which do not clarify which fixed effect of time is referred to, and because of this the reader can't understand which statistical test produced the reported p value.

Statistical Analysis section

a) Lines 195-96: "To compare global GMA...between groups Fisher's exact test was used." This does not tell the reader what time points the groups were compared using the Fisher’s test. Was it only baseline or endpoint, or at all time-points?

b) Lines 197-198 "A linear mixed model with a fixed effect for session, time point (before vs. after session) and intervention (IG vs. CG)...was used....."

i)"Mixed model" implies that in addition to the fixed effects, one or more random effects were included in the model, but no random effects are described in lines 197-98. Where the individual children included as a random effect? This should be stated in the initial description of the model.

ii) Later (Lines 237-40 discussed below) the text implies that the mixed model also included interaction terms for group*time-point and group*session and perhaps for group*time-point*session. All interaction terms that were included in the model should also be mentioned in the initial description of the model. This would greatly reduce confusion later in the text.
iii) Which measurement data was the mixed model used to analyze? Only the GMOS data? Or was it used to analyzed GMA data also?

c) Lines 200-201 "Exploratory post hoc analyses were performed to determine group differences for each measurement." This does not describe what test was used for post hoc analysis. Was the Fisher's test or the mixed model used for this, or a different kind of test?

Does "at each measurement" mean "at each time-point?"

Results: Secondary outcome

Line 237 "GMOS did not change over time (p=0.262)."

a) This sentence does not identify which test produced this p value. Do "over time" and the reported p value refer to a test of group*session interaction, or to a test of group*time point (before vs. after session) interaction? If there was a group*session term or a group*time-point term in the mixed model, they should be mentioned in lines 197-98.

b) Legend to Figure 2 presents the same p value (0.262) with the explanation "GMOS did not change significantly between baseline, during and the end of the intervention period in IR and CG (p=0.262)." ". between baseline, during and the end of the intervention" suggests a test of group*session interaction. The fact that this is the same p value as in Line 237 implies that "over time" in line 237 might mean "over the sessions," and that the p value given is from the same test of group*session interaction. But the next Line 238 states that "...no interaction of time*group could be observed (p=0.658)," and gives a different p value. So "time" in Line 238 seems to mean something different from "...between baseline, during and the end of the intervention" in Legend to Figure 2. What does "time" mean in Line 238? Does it mean "time-point" or "sessions" or something else?

Line 238-8 "Furthermore no differences between IG and CG (p=0.361)... could be observed."

a) Again, it is not clear what test produced this p value. Was it a test comparing IG and CG at baseline, at endpoint, or at all time-points? Was it a Fisher's exact test, was it part of the mixed model, or was it some other kind of test?

b) The next sentence says "Post hoc analysis showed a trend to higher values before (p=0.085) and after (p=0.075) the first session in CG compared to IG group." That statement seems to contradict the preceeding "...no differences between IG and CG...." Or does "no differences" mean "no statistically significant differences...?"

Line 238 ".no interaction of time*group could be observed (p=0.658)" But it is not clear whether "time" in this statement means "session" or "time point." Is it a test of significance of group*time-point interaction, or group*session interaction? Or does "time" in Line 238 means something else?

Line 239-40: "Post hoc analysis showed a trend to higher values before (p=0.85) and after (p=0.075) the first session in CG compared to IG group."
a) These seem to be p values for a group*time point*session interaction when session=1. If there is such an interaction term in the mixed model, it should be described in lines 197-98. Or was some other kind of statistical test used to compare GMOS values before and after the first session between groups?

b) This line 239-40 statement is contradicted by the labeling of Table 2, line 4 which presents the same p value (p=0.075) as a group effect at "the 12th assessment," not "after the first session."

Suggestions:

These many confusions about which interaction terms are in the mixed model, and which interaction terms go with which p values could be cleared up by a table that lists each term, including each interaction term, that is in in the model, and the values for beta, p, and 95% confidence interval that were estimated for each term.

Then lines 197-98 should be changed to include each of the interaction models, and to specify what the random effect(s) were.

It would also reduce confusion to replace "time" and "over time" at each point they occur in the text with "time-points" or "sessions," whichever fixed effect for time is being referred to. Or if "time" and "over time" mean some other kind of time, then that needs to be explained.

Summary, Results repeats the same confusions as lines 237-240 and should be corrected in the same way as that section. For the problem with each of the following statements read my comments in the preceeding section.

Lines 21-22: "The GMOS did not change significantly between baseline, during and the end of the intervention period (p=0.262) in the IG and the CG." Does this refer to a group*session interaction or a group*time-point interaction, or to some other test?

Line 23 "...no differences between IG and CG (p=262) ...could be observed." Do you mean "no statistically significant differences?" Or does this refer to some test comparing the two groups that was different from the test referred to in Lines 24-25.

Lines 24-25 "Post hoc analysis showed a trend toward higher values before (p=0.085) and after (p=0.075) the first session in CG compared to IG."

In addition to identifying which tests produced which p values, the Summary, Results section would be less confusing if it was preceded by a few sentences describing what statistical tests were performed in the analysis. Without this the reader does not know which tests the p values refer to.

Minor suggestions not for publication:

The ms. would benefit from a final light editing by an English language speaker, just to improve usage at a few points, e.g.

Lines 21-23 read "The GMOS (secondary outcome) did not change significantly
between baseline, during and the end of the intervention period (p=0.262) in the IG and the CG." Better usage would end the sentence with "in the IG or the CG."

Lines 197-198 "A linear mixed model with a fixed effect for session, time point (before vs. after session) and intervention (IG vs. CG)..." Better usage would be "A Linear mixed model with a fixed effects for ....."

Line 203 "A p-value of a<0.05 was considered statistical significant" should be "A p-value of p<0.05 was considered statistically significant."

Legend of Figure 2 "between baseline, during, and the end of the intervention period in IG and CG." Better usage would be "in IG or CG."

Summary, Line 8-9 "...or the control group (CG)?" perhaps it would be good to add a few words that describe what the control treatment was?

Figure 2 is much better organized now, but what is the meaning of the asterisk in the left hand margin? The Legend to Figure 2 has no item with an asterisk.