Author’s response to reviews

Title: The effect of a lifestyle intervention in obese pregnant women on change in gestational metabolic profiles: findings from the UK Pregnancies Better Eating and Activity Trial (UPBEAT) RCT.

Authors:

Debbie Lawlor (D.A.Lawlor@bristol.ac.uk)
Harriet Mills (harriet.mills@bristol.ac.uk)
Nashita Patel (nashita.r.patel@kcl.ac.uk)
Sara White (sara.white@kcl.ac.uk)
Dharmintra Pasupathy (dharmintra.pasupathy@kcl.ac.uk)
Annette Briley (annette.l.briley@kcl.ac.uk)
Diana Santos Ferreira (Diana.SantosFerreira@bristol.ac.uk)
Paul Seed (paul.seed@kcl.ac.uk)
Scott Nelson (Scott.Nelson@glasgow.ac.uk)
Naveed Sattar (Naveed.Sattar@glasgow.ac.uk)
Kate Tilling (Kate.Tilling@bristol.ac.uk)
Lucilla Poston (lucilla.poston@kcl.ac.uk)

Version: 2 Date: 06 Aug 2018

Author’s response to reviews:

Thank you for the opportunity to further revise this submission. Our response to all concerns are detailed below

Editorial concerns:

a.) Under the declaration 'Authors contributions', you have detailed the contributions from all authors except authors Sara White and Dharmintra Pasupathy; please clarify this.
RESPONSE: Apologies for this error. Dr Pasupathy contributed to the original design of UPBEAT and Dr White contributed to the design of the metabolomics study reported here; both gave critical comments on early drafts of the paper. This is now clear on the revised manuscript. [Pages 3-4, lines 65-74 of cleaned revision]

b.) Under the declaration 'Ethics and Consent', you've stated that you obtained approvals from participating local Research and Development (R and D) departments. Could you please provide us with those approval numbers/references?

RESPONSE: Apologies what we wrote was not correct. The local R&D departments provide assent for their centre to participate but not ethics approval and they do no provide reference numbers for that assent. We had already provided the reference for the ethics approval. The revised statement is as follows:

“Approvals were obtained from the UK research ethics committee (UK integrated research application system, reference 09/H0802/5) and assent for each centre to participate from local Research and Development (R and D) departments in each participating centres.” [Underline and bold is what has been added; Page 2, lines 27-28 of cleaned revision]

Please note that there are discrepancies between the 'clean' and 'tracked changes' versions of your manuscript, noted from the sections you point to in your responses to the reviewers comments; these are detailed below. Please amend this.

RESPONSE: We are really sorry for this and are not completely sure how it happened but we have now corrected everything, as follows.

Missing from clean version:
Reviewer 1, see responses 1 and 2

RESPONSE: We apologise one of the additions in response to comment 1 was omitted from the cleaned revised paper this is now added

“In total 62% of included women had all three measurements, 22% two measurements and 16% just one.” [Page 13, line 292-293 of cleaned revision]

And one line was omitted from the paragraph in response to point 2 from this reviewer. This has now been added:
“Sample processing was automated with a Gibson 215 liquid processor.” [Page 7, line 201 of cleaned revision]

Reviewer 3, see comment on multiple assessments

RESPONSE: The following is now included in the first section of the statistics analysis section

“We focus our discussion and interpretation of all results on the magnitudes of the point estimates (i.e. pregnancy change in metabolites or effect of the intervention) and their precision (i.e. 95% confidence intervals) as recommended by the American Statistics Society and others.[19-21] We explore the role of chance by providing exact p-values after controlling for multiple testing using the false discovery rate using the method of Benjamini and Hochberg.[22]” [Page 10, lines 222-227 of cleaned revision]

Reviewer 3, see comment on amendement of section in the Method page 11

RESPONSE: Again sorry that this did not transfer into the clean version. We have now made sure the paragraph is amended as stated in the response and as in the tracked version. The paragraph now reads:

We were keen to compare our findings in obese pregnant women to those in women not selected for being obese. As all of the participants in our study were selected for being obese, we were only able to do this indirectly, by searching the literature for other studies of similar metabolite profiles in general populations of pregnant women. We identified one previous study that examined cross-sectional differences using the same NMR metabolic profiles between women of reproductive age who were pregnant and those who were not (N = 4260 women; 322 of whom were pregnant). In addition to those cross-sectional analyses, longitudinal change in the metabolites were undertaken in a subgroup of women (N = 583) who were either pregnant at baseline and not at a follow-up assessment 6 years later, or were not-pregnant at baseline and were 6 years later.[4] That study also compared results separately by trimester of pregnancy. We compared the magnitude of longitudinal change, and differences by trimester, using summary data from that publication (specifically the results shown in figures 1 to 4 of that paper[4]) with our results to obtain some insight into whether pregnancy-related metabolic change differed between obese compared with non-obese women. [Pages 12-13, lines 271-285 of cleaned revision]

Reviewer 4, see comment on spline and smoothing methods (Method and Discussion text additions)
RESPONSE: We have now ensured that these comments are in the manuscript:

The following is the paragraph in the methods section

“Analysis assumptions and sensitivity analyses

Repeat metabolite assessments occurred at three time-points within a narrow range of gestational ages, such that there are gaps of up to 10 weeks with no (or very little) data between each of the measurements (sFigure 2). This means we had to use linear spline methods and could not explore smoothing methods or use fractional polynomials to determine the exact shape of metabolic trait change over pregnancy.[18] Furthermore, our main analyses assume that effect of the intervention is consistent between the first two measurements (~16-28 weeks) and the second two (~28 to 36 weeks). To test this assumption we modified the multilevel model to include the possibility that the magnitude of metabolite change might alter at 28-weeks, and visually compared the effect of the intervention for each trait between 16 to 28 weeks and 28 to 36 weeks. The linear spline method we have use assumes the model residuals are approximately Normally distributed, which may not be the case. There is evidence that estimates of population average change, such as those we present here are robust to non-Normality in the residuals (for example see reference[23]). However, to explore this further we have repeated our analyses of the effect of the intervention on differences in mean change of the metabolites using generalised estimating equations with robust standard errors and unstructured correlation matrices, which should be robust to non-normality and (some) misspecification of the working correlation matrices. Additional details of the linear spline model and its assumptions, including assumptions related to missing repeat measurements, are provided in supplementary text (sText).” [NOTE: this has some additional text related to further comments from reviewer 4 detailed below; Pages 11-12, lines 250-269 of cleaned revision]

The discussion text

“Whilst three repeat assessments of metabolites across gestation in a large RCT is unique, we were only able to fit linear spline multilevel models because metabolites were measured on just 3 occasions with very little variation in gestational age at measurement at those three times (see sFigure 2 and sText). This means that fitting non-linear models, for example using fractional polynomial or other ‘smoothing’ methods is not possible,[17, 18]]” was in fact in the clean revision 1 exactly as reported in the response to reviewer 4. It remains so in this second revision. [Page 16, lines 362-367 of cleaned revision]

Clean and Tracked versions read differently:
RESPONSE: We have now checked across both versions and ensured that the tracked and clean version are identical. We have uploaded the original submitted paper (i.e. prior to revision 1) together with revision 2 that includes ALL changes that have been made at both revisions.

Reviewer #4: A few minor issues:

1. In your response, you make a distinction between linear spline versus non-linear models. You mention a test of this assumption in the supplemental material. But there is no indication exactly how you tested this assumption, beyond the placement of the knot.

Did you test some hypothesis? If yes, was the power sufficiently high to detect a departure from a linear model that has substantive importance?

RESPONSE: We did not statistically test for a difference as this would over paramatise the model as noted in the supplementary material these analyses were undertaken by placing a knot point at 28 weeks of gestation and then including two interaction terms to test the effect of the intervention from 16 to 28 weeks and from 28 to 36 weeks. We then placed the results for each spline next to each other (supplementary Figure 3) so they can be compared visually. We have added the word ‘visually’ to the section of the methods that describes this:

“……and visually compared the effect of the intervention for each trait between 16 to 28 weeks and 28 to 36 weeks.” [Page 12, lines 258-260 of cleaned revision]

We had already (in both the original submission and revision 1) acknowledged in the discussion that this comparison would have limited statistical power and that remains unchanged in this second revision.

2. You characterize effect size in terms of standard deviations.

The standard deviation is not robust.

There is some possibility that there is a much stronger effect size if you use robust methods. What happens if you use, for example, a Winsorized standard deviation or some other robust measure of dispersion?

You cite [10] in the supplemental material to justify not worrying about non-normality. There are concerns about the design of their simulation study. Again, effect size might be more substantial in some sense than indicated in the paper.
RESPONSE: The standard deviation is only used to provide an appropriate scale comparison for the different metabolites, as they are all measured on such different scales. The p-values are derived from the analysis on the original scale and are therefore unaffected by choice of scaling multiplier. We have chosen the standard deviation of each metabolite in the control arm at baseline. We could have chosen a number of different multipliers (e.g. standard deviation in the entire group at baseline), and we could have looked at the effect of outliers (e.g. using winsorised sd as suggested). However, whilst the choice of scaling parameter will affect the numerical conclusions, it will not change the p-values for each hypothesis and is unlikely to change the interpretation of the results.

The reviewer states “There are concerns about the design of their simulation study.” in relation to our supplementary reference 10 but does not describe what these concerns are or who has raised them. That paper has been cited 6 times (2 by our group who were involved in the initial methods paper and 4 completely external) and none have noted concerns; there are no journal responses to the paper.

As the reviewer remains concerned about the possibility that we might find stronger effects with more robust methods we have now repeated the analyses using generalised estimating equations with robust standard errors and unstructured correlation matrices, which we consider to be one of the most appropriate methods with respect to being robust to non-normality and misspecification of the working correlation matrices. The results are identical to our original results (see supplementary Figure 4).

The following have been added to the revised paper:

Methods of main paper

“The linear spline method we have use assumes the model residuals are approximately Normally distributed, which may not be the case. There is evidence that estimates of population average change, such as those we present here are robust to non-Normality in the residuals (for example see reference[23]). However, to explore this further we have repeated our analyses of the effect of the intervention on differences in mean change of the metabolites using generalised estimating equations with robust standard errors and unstructured correlation matrices, which should be robust to non-normality and (some) misspecification of the working correlation matrices.” [Page 12, lines 260-267 of cleaned revision]

Results of main paper:
“Results for the effect of the intervention on change in metabolites were the same in our main multilevel linear spline models and the generalised estimating equations with robust standard errors suggesting that any non-Normality or outliers have not notably affected our results (sFigure 4).” [Page 15, lines 337-341 of cleaned revision]

Figure sFigure 4 has been added to the revised supplementary material – as this is the only change to supplementary material we have uploaded only the clean version of those materials.

Yours Sincerely,

Prof DA Lawlor on behalf of all co-authors