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: 1 Date: 25 Jun 2018

Author’s response to reviews:

Dear Diana Samuel,

Re our manuscript: "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." (BMED-D-18-00013).

Thank you for the opportunity to revise and resubmit this manuscript.
We would like to thank the reviewers whose comprehensive critiques have helped us to improve the paper. Below we give a point by point response to the reviewer’s comments.

Please note in these responses the page and line numbers refer to those on the ‘tracked’ (amended original manuscript) main and supplementary files

In addition to addressing the reviewers' comments, please also address the following editorial concern:

Editors comments

1. In accordance with our editorial policies, we require research articles to include a complete list of declarations. To that end, please include the declaration 'Consent for publication' (for further details, please see https://www.biomedcentral.com/getpublished/editorial-policies#consent+for+publication).

RESPONSE: Our manuscript does not include any individual participant/patient image, video or other person identifiable information and therefore a ‘consent for publication’ is not required.

2. Please also ensure that your revised manuscript conforms to the journal style, which can be found in the Instructions for Authors on the journal homepage.

RESPONSE: We have adhered to BMC Medicine Journal Style.

Reviewer #1:

The manuscript submitted by Mills and colleagues describes a NMR metabolomics study of 158 metabolic measures in blood in the UPBEAT RCT to assess changes in these measures during pregnancy in obese women. The team have also compared to a previous study in non-obese pregnancies. The manuscript is well written, defines the study accurately, does not over reach with its conclusions and is honest about the positives and limitations of the study performed. The work presented adds to our current knowledge of metabolic changes during pregnancy. There are a small number of recommended changes.

RESPONSE: We thank the reviewer for these supportive comments
1. Page 9, line 17. 1194 subjects has a minimum of one metabolic profile measure. It is unclear what is a metabolic measure is here and should be clarified - is this a sample for which NMR analysis was applied? If so, then a short discussion on how many subjects had two or three metabolic measures should be included and a comment on how assessing differences between subjects compared to assessing differences within the same subject influences the results.

RESPONSE: We apologise and agree with the reviewer that our original wording here was unclear. We have now revised this so that it reads:

“Of the 1194 from the remaining 6 centres 1158 (97%) had at least one set of nuclear magnetic resonance (NMR) metabolic measures together with complete data on ……” page 9, lines 189-190 in tracked revised manuscript

We have also added the following to the first paragraph of the results section:

“In total 62% of women had all three measurements, 22% two measurements and 16% just one.” page 13, lines 288-289 in tracked revised manuscript

In the methods section of the revised paper we now refer readers to further information about the missing at random assumption of multilevel models in Supplementary text. That sText is as follows:

“Multilevel models of change over time allow all participants with at least one measure to be included in analyses under the assumption that data are missing at random. The statistical term ‘missing at random (MAR)’ differs from missing completely at random (MCAR) in that it does not mean missingness is independent of all other characteristics. It means that conditional on the covariables included in the model (here age, parity, ethnicity, BMI and study centre) and the observed repeat measurements, the missing repeat measurements are not systematically different to those observed. This means that the effect of the intervention in those with some missing metabolic profile data is the same as in those with complete data at all three-time points conditional on the covariables included in the model. The MAR assumption would also be necessary if we restricted analyses to only those with all three repeat measurements (i.e. a complete case analysis). Whilst we cannot directly assess this assumption we feel that it is unlikely to be violated given most women had all three repeats, with just 16% having only one measure and that loss to follow-up in the trial was minimal and similar in both arms. The standard errors, and hence 95% confidence intervals, in these models take account of the greater random error of predicted levels at any time in those with just one or two of the repeat measurements.” Supplementary material, sText, pages 5-6, lines 156-171 of the tracked revised supplementary material
2. Venous blood was collected but no description of how these samples were processed was included - was serum or plasma prepared?

RESPONSE: We had partially described the processing in our original submission but agree with the reviewer that we had failed to describe whether we used serum or plasma and how these were prepared in fully. In the revised manuscript this full section now reads

“All blood samples were initially kept on dried ice, processed within 2-hours and then stored at -80°C until used for metabolic profiling. Metabolic analyses were undertaken on serum in two batches (March 2015 and August 2015), with samples from the three timepoints randomly distributed across these batches. Sample processing was automated with a Gibson 215 liquid processor. The complete process is illustrated in sFigure 1.” page 9, lines 198-203 in tracked revised manuscript

3. The first and third trimester samples were not fasted whereas the sample collected in the second trimester was fasted. It can be expected that fasting could influence the metabolic measures collected. The authors should comment/provide evidence on this and define whether they believe there will be metabolic changes between fasted and non-fasted blood samples and how this would influence the results described.

RESPONSE: We have added the following to the discussion section of the paper:

“Previous studies have used a mixture of fasted or non-fasted samples for metabolic measurements. In our study the first and third trimester samples were not fasted, whereas the second trimester one was. In previous analyses using the same NMR platform we have shown high levels of consistency between associations of these metabolites with cardiovascular diseases in studies using non-fasting and fasting samples.[14] Whilst we acknowledge the limited power of these analyses, the similarity between change from the first to second trimester measures and second to third trimester, also suggests that whether the samples are fasted or not does not have a marked impact on the results.” page 17-18, lines 396-403 in tracked revised manuscript

Reviewer #3: General comments: This is a well-designed and well-written study based on the combination of a well-executed randomized controlled trial of a lifestyle intervention in pregnancy in obese women and an earlier published observational longitudinal study. The study documents that the increases observed in obese pregnancies of specific metabolomics markers - which presumably reflect harmful processes - can be reduced by the intervention, and that these changes are pregnancy specific.
RESPONSE: We thank the reviewer for these supportive comments.

My main concern is the multiple assessments: is the report a fair representation of the evidence generated from these massive amounts of data? My understanding is that the results shown in the tables are only those, where tests met the conventional $p<0.05$ after they had been (somehow) adjusted for multiple comparisons (by controlling the false discovery rate); which is fine. However - and I may be wrong here as I am not a statistician - I think the authors could be more circumspect about how they present the results particularly in the abstract. For instance, expressing differential changes as multiples of standard deviation units may be misleading. To the uncritical reader they may give an exaggerated picture of what this study really shows, as the standard deviation units were (presumably) not adjusted for multiple comparisons (or were they?).

RESPONSE: Standard deviation units are frequently used to present change in multiple biomarkers so that comparisons over time between different markers can be made. In this paper it also allows us to compare with a previous study of women unselected for BMI. The p-values for these SD changes were adjusted for multiple comparisons, though we acknowledge that this was not clear in the original paper or supplementary statistical material. We have now made this clearer in the abstract (see below in response to specific points) and also in the main paper by moving the section that discusses this to the earlier part of the statistical analyses section and editing the words. This is now in the first section of the statistical methods in the main paper:

“We present exact p-values for all results but focus our discussion of the magnitudes of 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 10-11, lines 224-230 in tracked revised manuscript

Overall, I think this is an important study as it suggests that the intervention applied resulted in improved health profiles in obese pregnancy, as expressed by the assessed biomarkers. However, only further follow in mothers and children can document whether or not this will have implications in terms of improvements in hard health outcomes.

RESPONSE: We agree with the reviewer on this point and stated this in the original submission (this remains in this revision). However, we feel that the paper provides novel and unique information regardless of the current lack of follow-up.
Specific comments:

Abstract

Page 5, line 12: Please insert pregnant in "… (BMI) PREGNANT population …".
RESPONSE: This word has been inserted into the abstract as requested - Abstract, page 5, line 82 of revised tracked manuscript

Line 48: "… particles increasED .."  
RESPONSE: Thank you, as requested this typographical error has been corrected - Abstract, page 5, line 100 of revised tracked manuscript

Lines 44-48: Please reword this sentence, e.g., by using parentheses "All (extremely large, … , and very small) VLDL particles … , WHEREAS SPECIFIC (large, medium, and small) LDL particles …"
RESPONSE: These changes have been made exactly as requested - Abstract, page 5, lines 98-100 of revised tracked manuscript

Line 55: "larger than in unselected women from an independent, previously published study" - I think an abstract needs to contain the number of observations used, most readers will want to know that as basic information. Please somehow get this into the description.
RESPONSE: This section of the abstract has been revised accordingly so that it now reads as follows:

“… magnitudes of change across pregnancy in these obese women were 2-4 fold larger than in unselected women (N = 4260 in cross-sectional and 583 in longitudinal) from an independent, previously published, study.”  Abstract, page 6, lines 103-104 of revised tracked manuscript

Line 4-7. Please work more on the conclusion. Do the two statements, "Systemic metabolism …" and "with some evidence …", provide any real information? These statements seem a little too unspecific or general as a conclusion for this study - do you think they do enough justice to the large amounts of data that it generated?
RESPONSE: We have amended the conclusion of the abstract so that it now reads as follows:
“There are marked changes in multiple lipids and lipoproteins and more modest changes in other metabolites across pregnancy in obese women, with some evidence that this is more marked than in unselected (for BMI) pregnant women. The UPBEAT lifestyle intervention may contribute to a healthier metabolic profile in obese pregnant women, but our results require replication.”

Abstract, page 6, lines 107-112 of revised tracked manuscript

Also, the abstract should also somehow contain information indicating that the multiple comparison issue was addressed in this study. (Please also see my comment in "General comments" above.)

RESPONSE: We have added the following to the abstract:

“We focus primarily on the magnitudes and precision (95% confidence intervals) of estimates of change and differences between trial arms when describing our results. The role of chance is assessed with false discovery rate adjusted p-values.”

Abstract, page 5, lines 95-97 of revised tracked manuscript

Introduction

Page 6, line 1: I think that I would take "improved fatty acid profiles" out ("improved" in which sense? This can be debated.)

RESPONSE: We think this referred to a comment in the abstract (not introduction) and we have, as requested we have removed this clause. Deletion can be seen at Abstract, page 6, line 106 of the revised tracked manuscript

Page 8, line 1: You could help the reader if you added a few more words to explain "recent addition".

RESPONSE: We have changed this to the following:

“The addition of repeat (three occasions) gestational metabolic measurements to stored blood samples of participants from this RCT, which were added after completion of the main RCT had been published, provided a unique opportunity to determine the gestational metabolic profile in obese women and whether an intervention with known beneficial effects on diet, physical activity and adiposity influences this profile.”

Page 8, lines 155-160 of revised tracked manuscript
Methods

Page 11, 38-46: This sentence is too long, and as written the latter part ("… profiled between becoming pregnant after not being and vice versa …") is not understandable. Please amend.

RESPONSE: We have amended this section, including removing the term ‘vice-versa’ and adding more details of the study. It 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.” Page 12-13, lines 266-279 of revised tracked manuscript

Results

Page 13, lines 7-14. The characterization, "Indirect comparisons", seems a bit strange and maybe also be imprecise? Does it mean "non-randomized (i.e., in reality, potentially biased comparisons, as opposed to the comparisons within the randomized setting)? Furthermore, and more importantly, it is not clear where the reader can see the actual data underpinning the specific statements in this paragraph. Please refer to a specific table or figure.

RESPONSE: Indirect means that it is comparing results from one study (our RCT of obese pregnant women) for which we have individual participant data, to summary results from another published study. This compares to directly comparing metabolic profiled between obese and non-obese women within the same population, which we are unable to do. We have clarified this in the method section as detailed above (section added above Page 12-133, lines 266-279 of revised tracked manuscript). We have also added to that section of the methods the exact figures and tables in the previous manuscript (see above).
Line 23: I wondered if the word "stratified" is the correct term to describe this?

RESPONSE: We have removed the word stratified from that sentence so that it now reads as follows:

“Though statistical power is reduced in analyses comparing change between the first two and last two measurements we found that….“ Page 15, lines 331-332 of revised tracked manuscript

Discussion

Page 14, line 34-46. I am not sure that the fact that "there was similar loss to follow-up and proportions with metabolic profiles at each assessment in the two arms of the trial" necessarily can be taken to support "that any missing data on the metabolic profiles is missing at random". Couldn't they still be non-random in both arms? (Maybe equally 'non-random' across the two arms, we cannot know this.) Please consider this.

RESPONSE: Thankyou, see also response to Reviewer #1 point 1 also on this issue. We have clarified what the technical statistical term missing at random (MAR) means and how it differs from Missing Completely at Random (MCAR) in the supplementary text - referring to this text in the methods as follows:

“Additional details of the linear spline model and its assumptions, including assumptions related to missing repeat measurements, are provided in supplementary text (sText).”Page 12, lines 262-264 of revised tracked manuscript

In Supplementary text, the following has been added:

“Multilevel models of change over time allow all participants with at least one measure to be included in analyses under the assumption that data are missing at random. The statistical term ‘missing at random (MAR)’ differs from missing completely at random (MCAR) in that it does not mean missingness is independent of all other characteristics. It means that conditional on the covariables included in the model (here age, parity, ethnicity, BMI and study centre) and the observed repeat measurements, the missing repeat measurements are not systematically different to those observed. This means that the effect of the intervention in those with some missing metabolic profile data is the same as in those with complete data at all three-time points conditional on the covariables included in the model. The MAR assumption would also be necessary if we restricted analyses to only those will all three repeat measurements (i.e. a complete case analysis). Whilst we cannot directly assess this assumption we feel that it is unlikely to be violated given most women had all three repeats, with just 16% having only one measure and that loss to follow-up in the trial was minimal and similar in both arms. The standard errors, and hence 95% confidence intervals, in these models take account of the greater random error of predicted levels at any time in those with just one or two of the repeat
measurements.” Supplementary material, sText, pages 5-6, lines 156-171 of the tracked revised supplementary material

In the discussion we have also extended the discussion of this as follows:

“Given this is a well conducted RCT, and there was similar loss to follow-up and proportions with metabolic profiles at each assessment in the two arms of the trial, together with similar baseline characteristics between the overall arms of the trial and between the arms from centres that took blood samples, it is likely that this assumption is met.” Pages 16, lines 353-357 of revised tracked manuscript

Page 14, line 58: In my mind the multiple comparison issue is a more important issue for replication, than whether or not the model depicts the changes correctly or whether there may be some deviation from linearity. (We rarely have truly linear relationships in biology.)

RESPONSE: As noted earlier we have taken account of multiple testing across all of the analyses. Although we agree that this was not clear in our original submission. We are now clear throughout the paper that our interpretation of results is primarily based on the magnitude of point estimates and their precision (in line with the American Statistical Society and others), with the play of chance using false discovery adjusted p-values. We think it is important to discuss the linearity assumptions of our models and the statistical reviewer has requested additional information on this which we have provided (see below in response to Reviewer #4).

Table 2

In columns 2 and 3, what is give in brackets, it seems to be some ranges - is it 95% Cis, IQRs? And are these ranges somehow adjusted for multiple comparisons? This maybe stated in the Methods section, but it needs also to be completely clear from the table text. I cannot see it in the legend text or in the footnotes.

RESPONSE: We apologise for this omission and have now included in the column headings that these are 95% confidence interval – added to Revised Table 2.

The 95% confidence intervals are not adjusted for multiple comparisons. All methods that are used to adjust for the impact of multiple testing on the play of chance adjust the p-values only, which reflects the probability of the magnitude of the point estimate or a larger point estimate being found in a given sample, if the true ‘effect’ was null in the population to which inference is made. By contrast the 95% confidence intervals reflect the precision with which the point estimate is obtained in the sample.
Reviewer #4:

PLEASE NOTE WE HAVE MADE ONE RESPONSE TO THIS REVIEWER’S COMMENTS AS THE COMMENTS ARE ALL RELATED TO ALTERNATIVE METHODS FOR MODELLING THE REPEAT DATA IN OUR STUDY. THESE COMMENTS ARE AT THE END OF THE REVIEWER COMMENTS.

The statistical method used based on splines, is certainly an improvement on the usual linear model.

There are features of the method that might be improved and there are some technical concerns. Below are some details along with suggestions that might improve the data analysis portion of the study.

Splines are very popular among some statisticians. But compared to other smoothers that might be used, splines seem to be one of the least satisfactory techniques (e.g., Hardle, 1990; Wilcox, 2017).

Another concern is that the method used assumes that the error term has a normal distribution and a form of homoscedasticity is assumed. Non-normality can be a serious concern, even with a large sample size. Details are summarized in Wilcox (2017), which includes a range of techniques that might be used instead.

Included are robust smoothers.

These methods effectively deal with non-normality and heteroscedasticity.

Homoscedastic methods generally use an incorrect estimate of the standard error when there is heteroscedasticity.

In practical terms, confidence interval can be inaccurate regardless of how large the sample size might be.

There is a substantial literature for dealing with this concern and R functions for applying them are described in Wilcox (2017).

Diagnostics, such as testing the normality assumption, are known to be unsatisfactory.

Roughly when do such methods have enough power to detect situations where non-normality is a practical concern? In addition, even a slight departure from normality can create serious problems (e.g., Hampel et al., 1986; Huber and Ronchetti, 2009; Staudte and Sheather, 1990; Maronna, 2006).

Ultimately, modern robust might not make a difference. But I have seen many exceptions. So if these methods are ignored, some explanation is needed.
Means and standard deviations are not robust for reasons summarized in the books cited above. In Fig 2, for example, are similar results obtained using medians?

Distribution-free confidence intervals can be computed using the R functions in Wilcox (2017). The sample median trims all but one or two values. What happens when using the Tukey and McLaughlin (1963) method with a 20% trimmed mean? This method has been studied extensively. The R function trimci in Wilcox (2017) can be used.

In summary, suggest taking into account the many advances related to modern robust techniques. They can provide a deeper, more accurate and more nuanced understanding of data.


RESPONSE: We thank the reviewer for these comments. We realise that we may not have been sufficiently clear about the nature of our data. Because we have only three repeat measurements across pregnancy that were based on blood samples collected at antenatal assessments at three distinct time points across pregnancy we are only able to assess change with linear splines because there is very sparse data between these three time points. We have added to several parts of the revised manuscript to clarify this point as follows:

In the methods section of the paper:

“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. 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 compared the effect of the intervention for each trait between 16 to 28 weeks and 28 to 36 weeks. Additional details of the linear spline model and its assumptions, including assumptions related to missing repeat measurements, are provided in supplementary text (sText).” Pages 11-12, lines 252-264 of the revised tracked manuscript

In the supplementary material:

“Assumptions of the multilevel linear spline model

Due to the nature of the data collection, we cannot explore the pattern of change between the three timepoints, as the data in the intervening periods is sparse (see sFigure 2 below). Specifically, we had to use linear spline methods and could not explore smoothing methods, or use fractional polynomials,[3, 4] to determine the exact shape of metabolic trait change over pregnancy. The linear spline method we have had to use assumes the model residuals are approximately Normally distributed, which may not be the case with our data. However, 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[10]).” sText Page 5, lines 146-154, in the tracked changes revision

We have also added a new figure to supplementary material sFigure 2 page 38 supplementary material, which shows the issue that we face with virtually no data for periods of ~ 10 weeks between the first and second and second and third measurements. NOTE: the previous sFigure 2 is now sFigure 3 in the revised manuscript.

In the discussion the following has been added:

“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]" Page 16, lines 357-361 of the tracked changed revised manuscript.