PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Interpregnancy Weight Change and Adverse Pregnancy Outcomes: A Systematic Review and Meta-Analysis
AUTHORS	Oteng-Ntim, Mononen, Sofia; Eugene; Sawicki, Olga; Seed, Paul; Bick, Debra; Poston, Lucilla

VERSION 1 – REVIEW

REVIEWER	Jacqueline Wallace Rowett Institute, University of Aberdeen
REVIEW RETURNED	09-Aug-2017

GENERAL COMMENTS	The objective was to systematically review and conduct a meta-analysis on the effect of inter-pregnancy BMI change on selected pregnancy outcomes, namely LGA births, macrosomia, GDM and CS. Ten observational studies involving 910,951 women with singleton births at parity 0 and 1 were deemed suitable for the meta-analysis. The results were clear cut in that women who gained weight between conceptions were at higher risk of developing GDM, CS and LGA babies at the second pregnancy. Women who lost weight between pregnancies had a lower risk of GDM and LGA babies. This area of research is extremely familiar to me as I am the lead author of two of the studies used in this review. Although the review is well conducted I am concerned that the take home message states “Clinicians should aim to address weight change after birth of the first child in order to lower risk of adverse outcomes”. The implication is to encourage weight loss and I am a little puzzled as to why the authors did not mention or present analysis of the relationship between inter-pregnancy weight loss and the risk of primary or recurrent SGA. Both have been reported by myself (Wallace et al. 2014, 2016) and others (Bogaerts et al. 2013, Jain et al. 2013). As such remaining weight stable between consecutive conceptions could arguably be the more appropriate advice. Comments which require to be addressed are as follows: Table 1. Please clarify if parity and gender specific birth weight charts were used by all studies to define LGA. Similarly please clarify if elective or emergency caesarean section were combined, considered separately or only emergency used. Table 3. There are inaccuracies in the entries for Wallace 2014 and 2016. For 2014 the sample size is 12,740. For 2016 the sample size is 24,450 and the units of BMI change used to define categories were <-2, within -2 to +2 and >2 Figure 2 and 3. Significance of arrow head on CI of one line in each graph not defined. Limitations section: the effect of breastfeeding on postpartum weight retention is contentious at best so I modify the language here. It

	would be useful to segregate primary from recurrent risks at the second maternity and there is no mention of this distinction as far as I could determine.

REVIEWER	Christy Woolcott Dalhousie University, Canada
REVIEW RETURNED	19-Oct-2017

GENERAL COMMENTS	The authors have completed a meta-analysis addressing the association between interpregnancy change in body mass index (BMI) and perinatal outcomes (LGA, macrosomia, GDM, c-section). It is an important question because interpregnancy weight gain is one of the few potentially modifiable risk factors for adverse perinatal outcomes. To my knowledge, all relevant studies have been identified and included. A number of inconsistencies and potential errors in the data limit my confidence in the findings; these are listed below and I hope that they will be of use to the authors. Comments 1. The results reported in this manuscript have differences from those intended from the Prospero registration. The main difference being that not all of the outcomes have been represented in the manuscript. It is unclear whether no studies were found that reported results for the other outcomes (perinatal death, low birthweight/SGA, preterm birth, pre-eclampsia), or if these results are to be presented in a separate report. The reason for excluding these results should be stated. 2. The study by Villamor is stated in Table 1 as being a prospective cohort. I believe it was a retrospective cohort based on the Swedish Birth Register. 3. In table 1, several studies were noted to have the limitation, “low ethnic diversity”, which is not a limitation of the studies as it is simply a reflection of the populations in which these studies were conducted. Also, “no information on genetic factors”, “physical activity, diet” and “stress” were listed for many, but not all, studies where I do not believe any of the studies had this information. It is unclear to what genetic factors the authors are referring. 4. The statement made in the abstract, strengths and limitations, and manuscript text that 910,951 women were ‘enrolled’ in this meta-analysis is misleading and erroneously suggests that an individual-level pooled analysis was done. Technically 10 studies were included which represented data from 910,951 women. 5. Risk of bias: several studies did not achieve a point for adequacy of follow-up. I think this could be explained further - the missing data were in the exposure, correct? How much data needed to be missing to be considered inadequate? 6. The authors have defined ‘moderate’ and ‘substantial’ weight change categories very specifically in the text, but have not noted the exceptions that they have made for the five studies that do not report results based on these specific categories. 7. Data from the Whiteman study were included (e.g., Fig 4, the aOR of 1.41 was included as a moderate increase in BMI versus the reference): a) This OR is for the change from normal BMI in the 1st pregnancy to an obese BMI in the 2nd pregnancy, relative to women who stayed as a normal BMI for both pregnancies. The BMI change would necessarily be >5 kg/m2 and therefore, this aOR should be included in the other exposure category, “substantial”; b) it is unclear why the comparable data from the Getahun studies were not included: among women of normal BMI in the first pregnancy, relative to women remaining in the normal weight category, the risk

	of CS associated with becoming overweight (aOR=1.20) and with becoming obese (aOR=1.96); c) Suppl Fig 9, the Whiteman aOR of 1.41 has been included in both the 'moderate' and the 'substantial' categories. 8. "Strengths" section suggests that the authors have reduced heterogeneity by only including studies that have recognized and controlled for confounders. They have not justified this statement by assessing the effect of adjustment for potential confounders by comparing unadjusted and adjusted ORs. Furthermore, some studies have adjusted for gestational weight gain, which is questionable as a confounder (GWG in the first pregnancy would be responsible for a significant proportion of the interpregnancy weight change; GWG in the second pregnancy is a potential mediator and should not be adjusted for). 9. There are inconsistencies, most minor, throughout the text including: - The number of outcomes investigated. (Five, four, etc.) - Number of digits reported for aORs (e.g., 1, 2, and 3 within the abstract) - Definition of interpregnancy weight change (difference between first and second pregnancy of prepregnancy BMI versus BMI recorded at the first antenatal visit; not necessarily inconsistent, but difficult for readers to follow and, therefore, should be clarified). - Prospero registration number (abstract is incorrect) - The number of studies included in the meta-analysis (flowchart=10, but the two studies by Getahun et al. were not included in the main analyses presented in the body of the text). - The flowchart shows 8 studies excluded on the basis of 'first pregnancy complicated by other outcome' in but it is not part of the list of eligibility criteria in the text. Furthermore, it is unclear whether Wallace 2016, which was included and examines the recurrence risk of adverse outcomes in relation to interpregnancy weight change, should have been excluded on the basis of this criterion. - The abstract suggests that only population-based cohorts were included, but in reality, any type of observational study would have been eligible (including case-control studies and hospital-based cohorts). - It is stated that studies that included women with previous diabetes diagnoses were excluded. I believe they have excluded studies that were restricted to women with previous diabetes diagnoses. (Ehrlich's study was included and they adjust for GDM in the first pregnancy). - Some results shown in the forest plots shown in the main body of the text include those presented for only one of the BMI strata (e.g., Bogaerts data), but data from other studies that are restricted to only one of these strata are excluded (e.g., Getahun). Minor comments and suggestions: 1. Page 11, "slightly increased risk of c-section… aOR 1.05 (95% CI 0.89-1.23)" should be revised to "no association was observed between a decrease in a BMI and risk of c-section…". 2. In the implications section is stated: "…should be monitored after pregnancy to attempt to keep BMI change to a minimum". The word "change" should be replaced by "gain" or "increase". 3. Many statements have been repeated within the abstract and text. Some examples (not exhaustive): - Abstract: "with the study protocol registered a priori" is not necessary when the Registration number is included at the end of the abstract. - Specifics about the methods are included in the final

	paragraph of the introduction and then repeated in the methods section. - Methods: search and data extraction by two investigators is stated twice; sensitivity analysis description appears twice. 4. The word ‘effect’ in the stated objectives and elsewhere in the text should be replaced by ‘association’ to fit with the observational studies included in this review. 5. The search strategy includes “(c?esarian”) – spelling should be “(c?esarean)”? 6. Figure legends should be revised. For example, “change in interpregnancy weight with reference category and the risk of” could be better stated as “change in interpregnancy weight and risk of…, relative to the reference category [define]”.

REVIEWER	Ricardo Segurado University College Dublin, Ireland
REVIEW RETURNED	13-Nov-2017

GENERAL COMMENTS	Thank-you for the opportunity to review. This is a well-conducted and very well described study, which could benefit from some polishing but I have only minor points to address. However, I have had to mark it for major revision due to some major issues. The main one is that the protocol registered in PROSPERO has been substantially deviated from, without explanation in the manuscript. The paper is therefore at substantial risk of publication bias. I encourage the authors to include a section in their results indicating why they deviated from their protocol in the type of study, search strategy, exposure definition, population definition, primary and secondary outcomes, and lack of definition of the subgroup analyses or sensitivity analyses. The present manuscript does not deviate massively, but there are more differences than can be glossed over. I can believe that the main motivation may have been lack of relevant data or studies for the original plan, but if any post-hoc change occurred this should be stated. Other comments on the present manuscript: - From the abstract onwards the phrase "all women irrespective of BMI at first pregnancy is used" (e.g. results section, legend of figures). Since some of the included studies appear to have adjusted for pre-pregnancy BMI (Table 3), I don't think this statement is justified, or else the extracted data needs to be better explained. - I would avoid making claims here about macrosomia since there was only 1 study reporting this outcome (in Abstract, in table 8) Methods section: - The way in which the studies from Getahun were converted from Underweight/Normal/Overweight/Obese to Decrease/No Change/Moderate change/Substantial change needs to be clarified (figures 5, 6, 9, 10) - If the studies from the same country (with, I note, the same authors) are conducted in the same sample, this needs to be stated. It also impacts on the analysis, as one cannot include studies on the same participant as independent observations in a meta-analysis (e.g. Wallace 2014, 2016). - The authors need some assessment of publication bias for their

	primary outcomes, at least visually (funnel plots). Results section: - Figure 1, please clarify why the total 1335 studies found isn't the sum of those from each database (this is pre-duplicate removal?) - Table 3: it is not clear what is meant by genetic factors (specific candidate genes, family history?) - It needs to be clarified whether the classification of exposure as Underweight/Normal/Overweight/Obese (Getahun et al 2007a and b, Whiteman et al 2011a and b) was the reason these studies couldn't be used in the meta-analysis - All p-values stated as p=0.000 need to be changed to p<0.001 - The p-values for I2 statistics should probably be removed - I am not aware of any particularly sound inference on variance partitions, and they are likely to be anti-conservative - I suggest that you give the forest plots more clear headline titles, or up-front the legends, so that the outcome being presented can be seen at a glance - When looking at "Outcomes grouped by BMI before first pregnancy", the studies from Villamor et al, and Ehrlich et al are included, but these studies adjusted for pre-pregnancy BMI (Table 3). It needs to be stated that you either used an unadjusted odds ratio from these studies, or that they presented stratified results. Note that, even if they stratified results, they might have still adjusted for residual baseline BMI effects. - In the "Sensitivity analysis", with such a small number of studies, don't read much into the p-values. I suggest you don't discuss significance, only speak about the magnitude and direction of effect being similar (or not). - In the "Heterogeneity" section, the phrase "This value is thought to reflect the extent to which confidence intervals overlap each other". This understates how well we understand the statistic - which is well defined as a proportion of between-study variance. Discussion: - In paragraph 2 of the "Interpretation of major findings" section (end of page 13), there is an interesting mention of synergistic effects, which I believe are still an unknown factor. It would be good to say whether any (or how many) of the examined studies look at this, and whether the authors think it is a good idea (and maybe it should then be a recommendation in the Future Research section) - In paragraph 3 of the "Strengths" section, the authors state that adjusting for confounding means a reduced possibility that the results are due to chance, which is not strictly true - there is however a better possibility that the results are due to confounding or systematic bias - In the "Limitations" section, the relatively small sample size should be mentioned - although I note this is stated in Table 4. - In table 4, the first row on use of BMI "groups (underweight,

	obese)" the authors should respond to my comment above on the Getahun and Whiteman studies, and see if this row is needed. - I suggest that the authors make an effort, in the "Future Research" section to make recommendations for the improvement of, or improved reporting of, future observational studies on this topic. - e.g. consistency of endpoints or subgroup definitions, use of STROBE guidelines or ICHOM standards.

VERSION 1 – AUTHOR RESPONSE

Author's response to Reviewer(s)' Comments:

We would like to thank all of the reviewers for their time spent reviewing the manuscript and all the very useful comments, which identified areas that required improvement and helped clarify the manuscript. Below we have responded to each comment and indicated where in the manuscript each correction can be found.

Reviewer: 1
Reviewer name: Jacqueline Wallace

"Clinicians should aim to address weight change after birth of the first child in order to lower risk of adverse outcomes". The implication is to encourage weight loss and I am a little puzzled as to why the authors did not mention or present analysis of the relationship between inter-pregnancy weight loss and the risk of primary or recurrent SGA. Both have been reported by myself (Wallace et al. 2014, 2016) and others (Bogaerts et al. 2013, Jain et al. 2013). As such remaining weight stable between consecutive conceptions could arguably be the more appropriate advice."
Response: Thank you for your input, we have included SGA now as a fifth outcome and adjusted the advice appropriately to reflect this. Weight stability between first and second pregnancy is advised in order to reduce risk of adverse outcomes.

"Table 1. Please clarify if parity and gender specific birth weight charts were used by all studies to define LGA. Similarly please clarify if elective or emergency caesarean section were combined, considered separately or only emergency used."
Response: We have clarified that studies used similar birth weight charts for LGA and SGA under "outcome measures" and clarified that only emergency CS was considered in this study (table 1).

"Table 3. There are inaccuracies in the entries for Wallace 2014 and 2016. For 2014 the sample size is 12,740. For 2016 the sample size is 24,450 and the units of BMI change used to define categories were <-2, within -2 to +2 and >2"
Response: Thank you for this, the values have been corrected throughout (see table 3).

"Figure 2 and 3. Significance of arrow head on CI of one line in each graph not defined."
Response: The arrow head indicates the point at which the line has been truncated at the edge of the graph.This has been specified in the figure legend in appropriate figures (2 and 3).

"Limitations section: the effect of breastfeeding on postpartum weight retention is contentious at best so I modify the language here. It would be useful to segregate primary from recurrent risks at the second maternity and there is no mention of this distinction as far as I could determine."
Response: Thank you, the wording has been revised in limitations to reflect that breastfeeding is a contentious factor.

Reviewer: 2
Reviewer Name: Christy Woolcott

Comments
"1. The results reported in this manuscript have differences from those intended from the Prospero registration. The main difference being that not all of the outcomes have been represented in the manuscript. It is unclear whether no studies were found that reported results for the other outcomes (perinatal death, low birthweight/SGA, preterm birth, pre-eclampsia), or if these results are to be presented in a separate report. The reason for excluding these results should be stated."
Response: We have added discussion of our PROSPERO registration into our first paragraph of limitations to indicate why there are differences between registration and the manuscript. SGA has been added as an outcome into the results, the other outcomes were not included due to lack of relevant data, and this has been mentioned as a limitation.

"2. The study by Villamor is stated in Table 1 as being a prospective cohort. I believe it was a retrospective cohort based on the Swedish Birth Register."
Response: Thank you, this has been corrected in table 1.

"3. In table 1, several studies were noted to have the limitation, “low ethnic diversity”, which is not a limitation of the studies as it is simply a reflection of the populations in which these studies were conducted. Also, “no information on genetic factors”, “physical activity, diet” and “stress” were listed for many, but not all, studies where I do not believe any of the studies had this information. It is unclear to what genetic factors the authors are referring."
Response: Genetic factors have been changed to family history to avoid confusion, low ethnic diversity has been removed as a limitation. The other factors mentioned have been checked to be included as a limitation for all studies in table 1.

"4. The statement made in the abstract, strengths and limitations, and manuscript text that 910,951 women were ‘enrolled’ in this meta-analysis is misleading and erroneously suggests that an individual-level pooled analysis was done. Technically 10 studies were included which represented data from 910,951 women."
Response: This has been corrected in the abstract, strengths and limitations. Further, the data has been revised and corrected after the additional study added as part of SGA data.

"5. Risk of bias: several studies did not achieve a point for adequacy of follow-up. I think this could be explained further - the missing data were in the exposure, correct? How much data needed to be missing to be considered inadequate?"
Response: This has been specified in the appendix table 5 to reflect that any follow-up after birth was enough for a point, in addition to missing data that was adjusted for. Most research did not follow-up at all after birth and thus this was clarified.

"6. The authors have defined ‘moderate’ and ‘substantial’ weight change categories very specifically in the text, but have not noted the exceptions that they have made for the five studies that do not report results based on these specific categories."
Response:This has been addressed in “data collection and extraction” to help clarify. WHO classification was used for each study in order to convert from classes such as "overweight" (>25kg/m2) and "normal" (<25mg/m2) to changes in BMI units.

"7. Data from the Whiteman study were included (e.g., Fig 4, the aOR of 1.41 was included as a moderate increase in BMI versus the reference): a) This OR is for the change from normal BMI in the 1st pregnancy to an obese BMI in the 2nd pregnancy, relative to women who stayed as a normal BMI for both pregnancies. The BMI change would necessarily be >5 kg/m2 and therefore, this aOR should be included in the other exposure category, "substantial"; b) it is unclear why the comparable data from the Getahun studies were not included: among women of normal BMI in the first pregnancy, relative to women remaining in the normal weight category, the risk of CS associated with becoming overweight (aOR=1.20) and with becoming obese (aOR=1.96); c) Suppl Fig 9, the Whiteman aOR of 1.41 has been included in both the 'moderate' and the 'substantial' categories."
Response:
A) Thank you, this has been corrected in the graph by moving the OR to the correct category(substantial category) and removing from moderate increase in BMI.
B) Getahun data has been added, please refer to the updated figure 4.
C) Thank you, this has been corrected in suppl figure 9, Whiteman was only meant to be included in the substantial category.

"8. "Strengths" section suggests that the authors have reduced heterogeneity by only including studies that have recognized and controlled for confounders. They have not justified this statement by assessing the effect of adjustment for potential confounders by comparing unadjusted and adjusted ORs. Furthermore, some studies have adjusted for gestational weight gain, which is questionable as a confounder (GWG in the first pregnancy would be responsible for a significant proportion of the interpregnancy weight change; GWG in the second pregnancy is a potential mediator and should not be adjusted for)."
Response: This has been now discussed in limitations, as there was no unadjusted data to work with and it could not be calculated from the given data. GWG has also been discussed in paragraph 2 of the limitations and mentioned that it should not be adjusted for as it is a mediator in second pregnancy. .

"9. There are inconsistencies, most minor, throughout the text including:
- The number of outcomes investigated. (Five, four, etc.)"
Response: Corrected throughout text

"- Number of digits reported for aORs (e.g., 1, 2, and 3 within the abstract)"
Response: Corrected in abstract

"- Definition of interpregnancy weight change (difference between first and second pregnancy of prepregnancy BMI versus BMI recorded at the first antenatal visit; not necessarily inconsistent, but difficult for readers to follow and, therefore, should be clarified)" Response: Corrected in introduction

"- Prospero registration number (abstract is incorrect)"
Response: Corrected in abstract

"- The number of studies included in the meta-analysis (flowchart=10, but the two studies by Getahun et al. were not included in the main analyses presented in the body of the text). "
Response: Getahun has now been included in the main analyses and the flowchart has been corrected.

"- The flowchart shows 8 studies excluded on the basis of 'first pregnancy complicated by other outcome' in but it is not part of the list of eligibility criteria in the text. Furthermore, it is unclear whether Wallace 2016, which was included and examines the recurrence risk of adverse outcomes in relation to interpregnancy weight change, should have been excluded on the basis of this criterion."

Response: Corrected, was not meant to be part of eligibility criteria as Wallace would indeed then be excluded, thus flowchart has been revised and updated.

"- The abstract suggests that only population-based cohorts were included, but in reality, any type of observational study would have been eligible (including case-control studies and hospital-based cohorts)."
Response: We agree and therefore have corrected the abstract to reflect this, changing the wording from population-based cohorts to all observational studies.

"- It is stated that studies that included women with previous diabetes diagnoses were excluded. I believe they have excluded studies that were restricted to women with previous diabetes diagnoses. (Ehrlich's study was included and they adjust for GDM in the first pregnancy)."
Response: Corrected and specified to type II diabetes mellitus, as Ehrlichs study did adjust for GDM.

"- Some results shown in the forest plots shown in the main body of the text include those presented for only one of the BMI strata (e.g., Bogaerts data), but data from other studies that are restricted to only one of these strata are excluded (e.g., Getahun)."
Response: Thank you, Getahun has now been included in the meta-analysis (for example in figure 4).

Minor comments and suggestions:
"1. Page 11, "slightly increased risk of c-section… aOR 1.05 (95% CI 0.89-1.23)" should be revised to "no association was observed between a decrease in a BMI and risk of c-section…". "
Response: Corrected on page 11

"2. In the implications section is stated: "…should be monitored after pregnancy to attempt to keep BMI change to a minimum". The word "change" should be replaced by "gain" or "increase". "
Response: Corrected in implications section

"3. Many statements have been repeated within the abstract and text. Some examples (not exhaustive):
- Abstract: "with the study protocol registered a priori" is not necessary when the Registration number is included at the end of the abstract."
Response: Corrected in the abstract, a priori has been removed

"- Specifics about the methods are included in the final paragraph of the introduction and then repeated in the methods section." Response: The duplication has been corrected in the methods section.

"- Methods: search and data extraction by two investigators is stated twice; sensitivity analysis description appears twice."
Response: The duplication has been removed in methods.

"4. The word 'effect' in the stated objectives and elsewhere in the text should be replaced by 'association' to fit with the observational studies included in this review."
Response: Corrected throughout text

"5. The search strategy includes "(c?esarian") – spelling should be "(c?esarean)"?"
Response: The search strategy has been revised, please see figure 2.

"6. Figure legends should be revised. For example, "change in interpregnancy weight with reference category and the risk of" could be better stated as "change in interpregnancy weight and risk of…, relative to the reference category [define]"."

Response: Corrected in all figure legends, including supplementary file.

Reviewer: 3
Reviewer Name: Ricardo Segurado

"I encourage the authors to include a section in their results indicating why they deviated from their protocol in the type of study, search strategy, exposure definition, population definition, primary and secondary outcomes, and lack of definition of the subgroup analyses or sensitivity analyses. The present manuscript does not deviate massively, but there are more differences than can be glossed over. I can believe that the main motivation may have been lack of relevant data or studies for the original plan, but if any post-hoc change occurred this should be stated."
Response: Thank you, this has been discussed in the first paragraph of limitations as most of the outcomes had lack of relevant research. However, after review SGA has now been included as an additional outcome and the manuscript has been reviewed to reflect this.

Other comments on the present manuscript:

"- From the abstract onwards the phrase "all women irrespective of BMI at first pregnancy is used" (e.g. results section, legend of figures). Since some of the included studies appear to have adjusted for pre-pregnancy BMI (Table 3), I don't think this statement is justified, or else the extracted data needs to be better explained. "
Response: This has been corrected throughout the text

"- I would avoid making claims here about macrosomia since there was only 1 study reporting this outcome (in Abstract, in table 8) "
Response: This statement has been removed from the abstract

"Methods section:
- The way in which the studies from Getahun were converted from Underweight/Normal/Overweight/Obese to Decrease/No Change/Moderate change/Substantial change needs to be clarified (figures 5, 6, 9, 10) "
Response: This has been clarified in methods (see other reviewers comments above)

"- If the studies from the same country (with, I note, the same authors) are conducted in the same sample, this needs to be stated. It also impacts on the analysis, as one cannot include studies on the same participant as independent observations in a meta-analysis (e.g. Wallace 2014, 2016). "
Response: Thank you, this has been added into the study characteristics to clarify that the samples from same country/author were not the same sample.

"- The authors need some assessment of publication bias for their primary outcomes, at least visually (funnel plots)."
Response: We were unable to assess publication bias due to small amount of studies in each outcome (2-4) and have referred to papers suggesting at least 10 studies need to be in an outcome in order to complete funnel plots appropriately. We were not able to calculate the unadjusted odds ratios/original values for the data and thus assessment of publication bias was difficult in other ways. This has now been added and discussed in the limitations section.

"Results section:
- Figure 1, please clarify why the total 1335 studies found isn't the sum of those from each database (this is pre-duplicate removal?) "

Response: Thank you this is an error, corrected in figure 1

"- Table 3: it is not clear what is meant by genetic factors (specific candidate genes, family history?)"
Response: Corrected, specified as family history as no specific candidate genes were investigated

"- It needs to be clarified whether the classification of exposure as Underweight/Normal/Overweight/Obese (Getahun et al 2007a and b, Whiteman et al 2011a and b) was the reason these studies couldn't be used in the meta-analysis"
Response: Clarified in methods section and data has been reviewed (Getahun and Whiteman are used in the meta-analysis)

"- All p-values stated as p=0.000 need to be changed to p<0.001"
Response: Corrected throughout the text.

"- The p-values for I2 statistics should probably be removed - I am not aware of any particularly sound inference on variance partitions, and they are likely to be anti-conservative"
Response: These have been removed from the forest plots throughout the main text.

"- I suggest that you give the forest plots more clear headline titles, or up-front the legends, so that the outcome being presented can be seen at a glance"
Response: Headline titles have been added into each forest plot.

"- When looking at "Outcomes grouped by BMI before first pregnancy", the studies from Villamor et al, and Ehrlich et al are included, but these studies adjusted for pre-pregnancy BMI (Table 3). It needs to be stated that you either used an unadjusted odds ratio from these studies, or that they presented stratified results. Note that, even if they stratified results, they might have still adjusted for residual baseline BMI effects."
Response: Unadjusted odds ratios were not available nor could they be calculated from the data available, however this point has been added to limitations to discuss the effect of adjusted odds ratios and stratified results,

"- In the "Sensitivity analysis", with such a small number of studies, don't read much into the p-values. I suggest you don't discuss significance, only speak about the magnitude and direction of effect being similar (or not). "
Response: Wording has been revised in the sensitivity analysis section to remove discussion of the significance of the results.

"- In the "Heterogeneity" section, the phrase "This value is thought to reflect the extent to which confidence intervals overlap each other". This understates how well we understand the statistic - which is well defined as a proportion of between-study variance."
Response: The wording of this has also been revised in the heterogeneity section to better define heterogeneity as between-study variance.

"Discussion:
- In paragraph 2 of the "Interpretation of major findings" section (end of page 13), there is an interesting mention of synergistic effects, which I believe are still an unknown factor. It would be good to say whether any (or how many) of the examined studies look at this, and whether the authors think it is a good idea (and maybe it should then be a recommendation in the Future Research section)"
Response: This has been added into future research and specified that none of the examined studies looked at synergistic effects.

"- In paragraph 3 of the "Strengths" section, the authors state that adjusting for confounding means a reduced possibility that the results are due to chance, which is not strictly true - there is however a better possibility that the results are due to confounding or systematic bias"
Response: The wording of this has been revised to reflect bias instead of chance.

"- In the "Limitations" section, the relatively small sample size should be mentioned - although I note this is stated in Table 4. "
Response: This has been added into limitations.

"- In table 4, the first row on use of BMI "groups (underweight, obese)" the authors should respond to my comment above on the Getahun and Whiteman studies, and see if this row is needed."
Response: Thank you, this row has been kept as Getahun and Whiteman have now been included in the analysis.

"- I suggest that the authors make an effort, in the "Future Research" section to make recommendations for the improvement of, or improved reporting of, future observational studies on this topic. - e.g. consistency of endpoints or subgroup definitions, use of STROBE guidelines or ICHOM standards. "
Response: We agree, this has been added into future research.

VERSION 2 – REVIEW

REVIEWER	Ricardo Segurado University College Dublin, Ireland
REVIEW RETURNED	03-Jan-2018

GENERAL COMMENTS	I thank the authors for their consideration of my comments. While most of my comments have been addressed, the most substantial have been addressed superficially or not at all. Please address the remaining major comments below: 1) "I encourage the authors to include a section in their results indicating why they deviated from their protocol in the type of study, search strategy, exposure definition, population definition, primary and secondary outcomes, and lack of definition of the subgroup analyses or sensitivity analyses." The authors have explained that changes to the pre-specified outcomes are justified by exactly the reasons I hypothesised. However one must pay deeper attention to the registered protocol in terms of the other domains. Pre-registration is not a bureaucratic exercise, it serves a crucial scientific purpose. The difference in 'exposure' class definition and the inclusion of BMI change in baseline Normal BMI women need to be mentioned. 2) The way in which change in BMI groups (from e.g. the Whiteman & Getahun studies) was translated into a group of change in BMI is still not clear. As reviewer 2 also pointed out, one can deduce that a change from Normal to Obese must represent a "Substantial" increase in BMI. However, from Normal to Overweight may include varying proportions of "No change", "Moderate increase" and/or "Substantial increase". If these were discarded, that must be stated. Otherwise the authors should ensure that their methods are described in sufficient detail that a reader could replicate them.

	3) I am disappointed that the authors dismissed my question on sample overlap without the most rudimentary check. I was not satisfied with their assurance that the two Wallace (and other) studies were from independent samples of 12 thousand and 24 thousand from the same country, published within 2 years of each other. My own review of Wallace et al 2014 and 2016 revealed that the same cohort was used with identical selection criteria, except that the larger study extended the time period and probably included more hospitals within Aberdeen. Unless the authors have a personal communication from the original study authors that these are independent samples , in which case they must reference this communication in the manuscript, both Wallace studies cannot be included in the same meta-analysis. A similar issue arises due to highly probable overlap between the four (!) studies included that used the Missouri cohort (Getahun 2007a, 2007b, Whiteman 2011a, 2011b) particularly for the C-Section outcome. I suspect you will need to re-analyse some of the meta-analyses, and restructure Table 3 to appropriately indicate analyses of the same cohorts. 5) I must contradict the response on "Outcomes grouped by BMI before first pregnancy" Odds Ratios all being adjusted ORs. My reading of the Erlich study plainly revealed unadjusted odds ratios in Table 2, and I note that even the adjusted odds ratios were not adjusted for pre-pregnancy BMI; conversely, the Villamor study appears to report ORs simultaneously stratified by AND adjusted for baseline BMI. I believe these cannot be combined. This meta-analysis needs to be reviewed in light of this fundamental difference - I don't think they can be combined. I noticed an error in the pages for the Getahun 2007a reference, I suggest the references be proof-read. Finally, a simple comment on the funnel plot response. While it is true that small numbers of studies do not give a particularly useful funnel plot, the same applies to meta-analysis itself and a quick search would reveal influential statisticians recommending a minimum number of 5 or of 9 or of 10 studies for a random effects meta-analysis. I think the authors are in a bind here, as what is good for the goose should also be good for the gander. I am willing to let it go as one can inspect the forest plots to gauge the same information, with a little mental gymnastics. All other comments were adequately addressed.

REVIEWER	Jacqueline Wallace Rowett Institute, University of Aberdeen, UK
REVIEW RETURNED	09-Jan-2018

GENERAL COMMENTS	Thank you for the opportunity to review this revised manuscript. I am pleased that the authors have now included SGA as one of their outcomes as I firmly believe the significant relationship between inter-pregnancy weight loss and a heightened risk of SGA birth to be both important and potentially overlooked when providing weight management advice during the childbearing period. Although the authors have now included the SGA data in their revised manuscript results section I would still like to see a more balanced consideration of the data in the discussion section – please address. Furthermore in their rebuttal they state “weight stability between first and second pregnancy is advised in order to reduce risk of adverse outcomes” but this is not expressly stated in the abstract and I believe it should

	be. My other comments on the first version of the manuscript have largely been addressed. I have picked up a few minor issues in the current version which the authors should also consider and address. Page 4 line 30: “Five” not Four. Page 5 line 5: While the reference weight/BMI stable category in Wallace et al 2014 was as stated i.e. between -1 and 1 units, it was wider in the second Wallace 2016 paper i.e. between -2 and 2 units. This was considered more stringent particularly with regard to defining inter-pregnancy weight loss (i.e. >2 units loss). In both instances the weight and height were measured in clinic and weight was adjusted to a standard stage of gestation for all women in the cohort. I consider this worthy of comment. In contrast in the Whiteman study a women could have experienced an inter-pregnancy BMI change of up to 5 units but stayed within the same BMI category and be categorised as weight stable. The text here should more accurately reflect the actual reference categories used. Similarly on Page 6 line 22, text should perhaps read “defined as women who remained within their original BMI category or their BMI changed by up to 2 units in either direction”. In Table 3 the BMI unit changes are wrong for Wallace et al 2014 (above text is correct). In Table 3 there are also inconsistences with regard to the limitations section of the table, particularly with respect to whether or not maternal weight measured in late pregnancy. Page 6 line 5: “eleven” not ten Page 6 line 12: “Seven” not S. The authors have included one paper (ref 16) that studied macrosomia (>4000g) but seem to have ignored the fact that the same manuscript reports data for low birth weight (<2500g). As inter-pregnancy weight loss doubled the rate of low birth weight this is also worthy of mention now that SGA has been included. This should be included on Page 9, ~ line 10-13. Page 9 line 30: Outcomes grouped by BMI before first pregnancy – SGA also needs to be included here (and as supplementary information in Table 2 with additional figure) as data is available. Importantly Wallace et al. 2014 shows that the relationship between inter-pregnancy weight loss and the primary risk of SGA at the second delivery is similarly strong in women who had a BMI<25 versus those with a BMI >25 at first pregnancy. Page 10 lines 42-44: This statement is at odds with the data in terms of weight loss and SGA and needs to be revised. The subsequent line only focuses on gain and no mention is made of loss. Overweight/obese women are at risk of substantial weight loss as well as gain between pregnancies. Although the latter is more common the consequences of weight loss in terms of increased risk of SGA are serious. The authors may wish to consider our recent paper https://www.ncbi.nlm.nih.gov/pubmed/28628636 Page 11 line 31, limitations: I did not realise that the PROSPERO registration included preterm birth and pre-eclampsia. Several of the studies already cited in this review include both outcomes and in my opinion could have been examined herein. Page 12 line 48: Do the authors mean “postpartum” rather than gestational weight loss here? Figure legends for 2-5 require to reflect variation in width of reference category as outlined above.

VERSION 2 – AUTHOR RESPONSE

We would like to sincerely thank the reviewers for reviewing the manuscript again and sending their helpful comments. Below we have indicated our responses to each comment and outlined the changes made to the manuscript.
Reviewer: 3
Reviewer Name: Ricardo Segurado

“1) "I encourage the authors to include a section in their results indicating why they deviated from their protocol in the type of study, search strategy, exposure definition, population definition, primary and secondary outcomes, and lack of definition of the subgroup analyses or sensitivity analyses."
The authors have explained that changes to the pre-specified outcomes are justified by exactly the reasons I hypothesised. However one must pay deeper attention to the registered protocol in terms of the other domains. Pre-registration is not a bureaucratic exercise, it serves a crucial scientific purpose. The difference in 'exposure' class definition and the inclusion of BMI change in baseline Normal BMI women need to be mentioned.
Response: Regarding the deviation from our protocol: the type of study was similar in the protocol and manuscript (observational studies). Search strategy was similar as the one in the registration except we do note that the outcomes differ and this is due to funding limitations (this research was internally funded). We understand that in the manuscript the exposure/population definition included women of all weight and not just obese/overweight – this was due to many studies including results with “all women” regardless of initial BMI (which is more representative of the population), and to not overlook the importance of weight change in normal weight women. We have mentioned this point in the limitations section. Subgroup analyses has been defined in report, as at time of protocol we did not have details of subgroup/sensitivity analyses. We take seed from the reviewers encouragement.
2) The way in which change in BMI groups (from e.g. the Whiteman & Getahun studies) was translated into a group of change in BMI is still not clear. As reviewer 2 also pointed out, one can deduce that a change from Normal to Obese must represent a "Substantial" increase in BMI. However, from Normal to Overweight may include varying proportions of "No change", "Moderate increase" and/or "Substantial increase". If these were discarded, that must be stated. Otherwise the authors should ensure that their methods are described in sufficient detail that a reader could replicate them.
Response: To study whether association between change in body weight and adverse outcomes differed, study groups were classified as “substantial increase in BMI”, “moderate increase in BMI” and “decrease in BMI”. These groups were defined as BMI increase of more than 3 units (substantial increase), BMI increase between 1 and 3 units (moderate increase) and BMI decrease more than 1 units (decrease). We have now specified the reference category as also including women who remained within their BMI category or their BMI changed by up to 2 units in either direction. In the case of Whiteman and Getahun studies, we have clarified the graphs so that we only used data which represented a substantial increase in BMI, (normal to obese) or weight loss (normal to underweight) in the appropriate places. The other data was not discarded but used in the subgroup analysis (<25 and >25) as appropriate. We have explained this in more detail in the methods section, page 4: “In studies that reported results based on WHO classification, women who changed from normal weight to underweight were considered as part of the BMI decrease category, and weight change from normal to obese represented a substantial increase in BMI. These studies were used as part of subgroup analyses (initial BMI >25 “overweight/obese” or BMI < 25 “normal”) and converted into substantial (normal to obese), moderate (normal to overweight) and decrease in BMI groups (normal to underweight) respectively.”

3) I am disappointed that the authors dismissed my question on sample overlap without the most rudimentary check. I was not satisfied with their assurance that the two Wallace (and other) studies were from independent samples of 12 thousand and 24 thousand from the same country, published within 2 years of each other. My own review of Wallace et al 2014 and 2016 revealed that the same cohort was used with identical selection criteria, except that the larger study extended the time period

and probably included more hospitals within Aberdeen. Unless the authors have a personal communication from the original study authors that these are independent samples , in which case they must reference this communication in the manuscript, both Wallace studies cannot be included in the same meta-analysis. A similar issue arises due to highly probable overlap between the four (!) studies included that used the Missouri cohort (Getahun 2007a, 2007b, Whiteman 2011a, 2011b) particularly for the C-Section outcome. I suspect you will need to re-analyse some of the meta-analyses, and restructure Table 3 to appropriately indicate analyses of the same cohorts.
Response: We apologise for dismissing this point without proper explanation – we have analysed this again by removing Wallace 2014 from LGA/SGA results as well as removing both Wallace 2014 and Whiteman in the CS outcome to avoid overlap. These results have now been included in the supplementary file to show that the direction of effect remained the same and thus did not change the conclusions. We would like to make note that we feel it was not appropriate to contact J Wallace as she is the first reviewer of this manuscript and therefore we have instead completed a sensitivity analysis included in supplementary file (see also added section at end of "outcomes").

5) I must contradict the response on "Outcomes grouped by BMI before first pregnancy" Odds Ratios all being adjusted ORs. My reading of the Erlich study plainly revealed unadjusted odds ratios in Table 2, and I note that even the adjusted odds ratios were not adjusted for pre-pregnancy BMI; conversely, the Villamor study appears to report ORs simultaneously stratified by AND adjusted for baseline BMI. I believe these cannot be combined. This meta-analysis needs to be reviewed in light of this fundamental difference - I don't think they can be combined.
Response: We only used adjusted odds ratios as many of the studies did not give us raw data and we were therefore not able to calculate unadjusted odds ratios. The differences in adjustment has been noted in table 3. The timing between the recording of baseline BMI in Villamor and the BMI at first pregnancy in the Ehrlich differed only by a week, further, Ehrlich (although did not adjust for pre-pregnancy BMI) adjusted for the time at which BMI was recorded in first pregnancy (16.7 weeks compared with 15 weeks in Villamor study). As there are so few studies included in the meta-analysis is it difficult to find studies that have adjusted for the same confounding factors, thus these points have been included in table 3.

I noticed an error in the pages for the Getahun 2007a reference, I suggest the references be proof-read.
Response: Thank you, this has been corrected.

"Finally, a simple comment on the funnel plot response. While it is true that small numbers of studies do not give a particularly useful funnel plot, the same applies to meta-analysis itself and a quick search would reveal influential statisticians recommending a minimum number of 5 or of 9 or of 10 studies for a random effects meta-analysis. I think the authors are in a bind here, as what is good for the goose should also be good for the gander. I am willing to let it go as one can inspect the forest plots to gauge the same information, with a little mental gymnastics. "
Response: Thank you for this comment, we agree that the lack of relevant studies is a limitation in this meta-analysis and have mentioned it accordingly.

Reviewer: 1
Reviewer Name: Jacqueline Wallace

"Although the authors have now included the SGA data in their revised manuscript results section I would still like to see a more balanced consideration of the data in the discussion section – please address." Response: Thank you, we have added more depth of discussion regarding SGA in the discussion section as well as referring to the paper you kindly linked to us.

“Furthermore in their rebuttal they state “weight stability between first and second pregnancy is advised in order to reduce risk of adverse outcomes” but this is not expressly stated in the abstract and I believe it should be.” Response: Thank you, we have now expressly stated this in the abstract.

My other comments on the first version of the manuscript have largely been addressed. I have picked up a few minor issues in the current version which the authors should also consider and address.
“Page 4 line 30: “Five” not Four. “
“Page 6 line 5: “eleven” not ten”
“Page 6 line 12: “Seven” not S.”
Response: Thank you, these mistakes have been corrected.

“Page 5 line 5: While the reference weight/BMI stable category in Wallace et al 2014 was as stated i.e. between -1 and 1 units, it was wider in the second Wallace 2016 paper i.e. between -2 and 2 units. This was considered more stringent particularly with regard to defining inter-pregnancy weight loss (i.e. >2 units loss). In both instances the weight and height were measured in clinic and weight was adjusted to a standard stage of gestation for all women in the cohort. I consider this worthy of comment. In contrast in the Whiteman study a women could have experienced an inter-pregnancy BMI change of up to 5 units but stayed within the same BMI category and be categorised as weight stable. The text here should more accurately reflect the actual reference categories used. “ “Figure legends for 2-5 require to reflect variation in width of reference category as outlined above.”
Response: Thank you, the text now reflects the reference category as being women who remained in the BMI category or changed by +- 2 units. This has also been changed in the figure legends.

“Similarly on Page 6 line 22, text should perhaps read “defined as women who remained within their original BMI category or their BMI changed by up to 2 units in either direction”. “ Response: Thank you, this has been corrected.

“In Table 3 the BMI unit changes are wrong for Wallace et al 2014 (above text is correct).” Response: This has been corrected in table 3.
In Table 3 there are also inconsistences with regard to the limitations section of the table, particularly with respect to whether or not maternal weight measured in late pregnancy. “
Response: Thank you for pointing this out, table 3 has been reviewed to specify if studies used pre-pregnancy BMI or baseline BMI at first pregnancy to avoid this inconsistency.
“ The authors have included one paper (ref 16) that studied macrosomia (>4000g) but seem to have ignored the fact that the same manuscript reports data for low birth weight (<2500g). As inter-pregnancy weight loss doubled the rate of low birth weight this is also worthy of mention now that SGA has been included. This should be included on Page 9, ~ line 10-13. “ Response: Thank you, this has been included on page 9.

“Page 9 line 30: Outcomes grouped by BMI before first pregnancy – SGA also needs to be included here (and as supplementary information in Table 2 with additional figure) as data is available. Importantly Wallace et al. 2014 shows that the relationship between inter-pregnancy weight loss and the primary risk of SGA at the second delivery is similarly strong in women who had a BMI<25 versus those with a BMI >25 at first pregnancy. “
Response: Jain et al, Cheng et al did not stratify data with BMI <25 and BMI >25 as did Wallace in 2014 and thus a supplementary figure was not completed. We have, however, added the information about Wallace et al 2014 and SGA in females > and <25 under the “Outcomes grouped by BMI before first pregnancy” section (see page 10).

“Page 10 lines 42-44: This statement is at odds with the data in terms of weight loss and SGA and needs to be revised. The subsequent line only focuses on gain and no mention is made of loss. Overweight/obese women are at risk of substantial weight loss as well as gain between pregnancies.

Although the latter is more common the consequences of weight loss in terms of increased risk of SGA are serious. The authors may wish to consider our recent paper https://www.ncbi.nlm.nih.gov/pubmed/28628636 “
Response: Thank you, we have updated this statement as well as added your recent paper into the discussion (reference 38).

“Page 11 line 31, limitations: I did not realise that the PROSPERO registration included preterm birth and pre-eclampsia. Several of the studies already cited in this review include both outcomes and in my opinion could have been examined herein. “
Response: As discussed above, the PROSPERO registration does unfortunately deviate from the manuscript in regards to the outcomes. Some of the data was poor quality and funding was lacking to complete all outcomes as originally intended. More in-depth discussion about the deviations from protocol has been included in the limitations section.

“Page 12 line 48: Do the authors mean “postpartum” rather than gestational weight loss here? “
Response: Thank you, this has been corrected.

VERSION 3 – REVIEW

REVIEWER	Jacqueline Wallace Rowett Institute University of Aberdeen UK
REVIEW RETURNED	12-Mar-2018

GENERAL COMMENTS	By addressing the reviewers detailed comments the manuscript is now balanced and much improved from the original version. I have one minor grammatical suggestion for further improvement which the authors may wish to consider and two comments, but irrespective I consider the manuscript should be accepted. Reviewer 3 comment 3 and author response: I am not convinced that either party fully understand the difference in focus between the Wallace et al 2014 paper and the 2016 one. The reviewer correctly identifies that the selection criteria were identical in terms of data extracted and that the more recent paper extended the period of data extraction and hence resulted in a doubling of the numbers of women involved – all delivered in a single hospital. However in terms of logistic regression analysis the 2014 paper mainly deals with the relationship between inter-pregnancy weight change and the primary complication incidence at the second pregnancy for women without that complication at first pregnancy while the 2016 is concerned with recurrent pregnancy complications only. This important distinction seems to have been missed and importantly a women who had LGA (or any of the other complications systematically reviewed) at the second pregnancy only in the 2014 paper would not have been considered in the 2016 paper as it examined the relationship between inter-pregnancy weight change in women who had LGA (or any of the other complications studies) in both pregnancies. Page 12 of tracked version line 6-8: all three Wallace papers highlight the association between inter-pregnancy weight loss and the increased risk of low placental weight and SGA birth. Page 12 of tracked version: second line from bottom suggest “which” enhances ….. provides (plural)

REVIEWER	Ricardo Segurado University College Dublin, Ireland
REVIEW RETURNED	16-Mar-2018

GENERAL COMMENTS	I sincerely thank the authors for engaging with all my queries and concerns. I believe the way you have addressed sample overlap is still incorrect. It is at least clear in this version that you have included (likely) overlapping studies. At this point I must leave it to the editor to decide whether to publish a technically incorrect analysis, which might be misleading if the manuscript isn't read in detail. I am still recommending on a point of principle that you fully address this issue. There is now a discrepancy in that Appendix 2 appears to have addressed this issue, but I am unclear what is presented in the "Outcomes" section in the body of the manuscript, and the "Outcomes grouped by BMI..." section and Appendix 3 have certainly not addressed the issue. I disagree with calling this a "sensitivity analysis", as it is in fact the correct analysis and should be the main analysis, although I believe you only refer to it thus in the response, and not in the revised manuscript? To re-iterate: it is a fundamental assumption of your meta-analytic statistics that your samples do not overlap. This is likely to bias the meta-analysis results, reduce your standard errors inappropriately, and decrease (i.e. improve) the estimated heterogeneity of the study-level effects inappropriately.

VERSION 3 – AUTHOR RESPONSE

We would like to sincerely thank the reviewers for their time and input regarding this manuscript. We have addressed the comments and hope that our responses address the feedback appropriately.

Reviewer: 1

"Reviewer 3 comment 3 and author response: I am not convinced that either party fully understand the difference in focus between the Wallace et al 2014 paper and the 2016 one. The reviewer correctly identifies that the selection criteria were identical in terms of data extracted and that the more recent paper extended the period of data extraction and hence resulted in a doubling of the numbers of women involved – all delivered in a single hospital. However in terms of logistic regression analysis the 2014 paper mainly deals with the relationship between inter-pregnancy weight change and the primary complication incidence at the second pregnancy for women without that complication at first pregnancy while the 2016 is concerned with recurrent pregnancy complications only. This important distinction seems to have been missed and importantly a women who had LGA (or any of the other complications systematically reviewed) at the second pregnancy only in the 2014 paper would not have been considered in the 2016 paper as it examined the relationship between inter-pregnancy weight change in women who had LGA (or any of the other complications studies) in both pregnancies. " Thank you for this clarification, we have addressed the comments of reviewer 3 to alleviate confusion by re-analysing the data and only including one Wallace paper in each subgroup to prevent overlap of the cohorts.

"Page 12 of tracked version line 6-8: all three Wallace papers highlight the association between inter-pregnancy weight loss and the increased risk of low placental weight and SGA birth. "

Thank you, this has been amended on page 12 to include reflect that all three Wallace papers showed this association.

"Page 12 of tracked version: second line from bottom suggest “which” enhances ….. provides (plural) "

This grammatical error has been corrected.

Reviewer: 3

"I sincerely thank the authors for engaging with all my queries and concerns. I believe the way you have addressed sample overlap is still incorrect. It is at least clear in this version that you have included (likely) overlapping studies. At this point I must leave it to the editor to decide whether to publish a technically incorrect analysis, which might be misleading if the manuscript isn't read in detail. I am still recommending on a point of principle that you fully address this issue. There is now a discrepancy in that Appendix 2 appears to have addressed this issue, but I am unclear what is presented in the "Outcomes" section in the body of the manuscript, and the "Outcomes grouped by BMI..." section and Appendix 3 have certainly not addressed the issue. I disagree with calling this a "sensitivity analysis", as it is in fact the correct analysis and should be the main analysis, although I believe you only refer to it thus in the response, and not in the revised manuscript? To re-iterate: it is a fundamental assumption of your meta-analytic statistics that your samples do not overlap. This is likely to bias the meta-analysis results, reduce your standard errors inappropriately, and decrease (i.e. improve) the estimated heterogeneity of the study-level effects inappropriately. "

Thank you, we have decided to re-do the analysis to remove any possible overlap of cohorts in the manuscript. This included re-analysing the main groups and sub-groups throughout the manuscript (specifically making sure that the two Wallace papers and the papers from the Missouri cohort were not analysed in the same group). The results have been updated accordingly (removing these papers did not change the significance or direction of effect of the results). Hopefully this addresses the concerns and alleviates any confusion!

VERSION 4 – REVIEW

REVIEWER	Ricardo Segurado UCD, Ireland
REVIEW RETURNED	11-Apr-2018

GENERAL COMMENTS	Compliments to the authors and my co-reviewers for assisting in getting this over the line. The study is clearer and more sound for it, despite the smaller number of studies within some of the meta-analyses.