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)	Alterations in maternally-perceived fetal movement and their association with late stillbirth – findings from the Midland and North of England Stillbirth Case-Control Study
AUTHORS	Heazell, Alexander; Budd, Jayne; Li, Minglan; Cronin, Robin; Bradford, Billie; McCowan, Lesley; Mitchell, Edwin; Stacey, Tomasina; Martin, Bill; Roberts, Devender; Thompron, John

VERSION 1 – REVIEW

REVIEWER	Brita Askeland Winje Norwegian Institute of Public Health, Oslo, Norway
REVIEW RETURNED	06-Nov-2017

| GENERAL COMMENTS | Thank you for the opportunity to review the manuscript entitled "Alterations in maternally-percieved fetal movement and their association with late stillbirth – findings from the Midland and North of England Stillbirth Case-Control Study"

Unfortunately I have concerns over the methods applied and the manuscript needs substantail revision before publication.

The manuscript is reporting fetal activity in pregnancies from women who have experienced stillbirth (cases) compared with gestational age matched women with ongoing pregnancies (controls). The choice of the case-control design is well justified by the authors and risk of recall bias is discussed in the paper. The research team is strong with long-term commitment to fetal movement research. I will commend the team for including additional characteristics of fetal movement such as strengths of the movement and hiccups and not only the frequency. This has been far less studied and may provide important input on how to inform women on when to seek medical attention.

However, I have concerns about the manuscript in its current form:

1. I am especially concerned about the combined frequency/strength variable that was constructed from Supplemental Table 1. This is not clear to me. As far as I understand it, Supplemental Table 1 was formed through a regression that essentially interacted categorical frequency with categorical strength. Similar odds ratios (e.g. 0.2-0.5, 2.0-3.0) were grouped together, and then the respective frequency/strength combinations were grouped together to form a more compact variable, which was then run in a secondary regression analysis (the main analysis for this paper).

If correctly understood, I am not sure this is an appropriate analysis strategy. When you form this more compact variable, you already |

know the results of the regression. You have already "peeked behind the curtain", so to speak, and thus this new variable will be inherently biased.

The hypothesis tests that have been performed are appropriate for testing a-priori hypotheses. To my understanding, Supplemental Table 1 is an a-priori hypothesis. The compact/combination variable is not an a-priori hypothesis, and as such, it is not appropriate to analyse this variable.

I see three available options:
a) Use supplemental Table 1 (or something similar) as your results (i.e. interact, but don't make it compact/grouped)
b) Run a regression without interacting frequency/strength (i.e. just run it as two variables that are each weaker/no change/stronger)
c) Keep your compact/grouped variable, but validate it in a new independent cohort (in which this compact/grouped variable would be able to be tested in an a-priori manner without any bias issues)

If I have understood the methods correctly, the manuscript needs to be re-analyzed before it is acceptable for publication.

2. It appears from the Methods that the fetal activity data were added to a model that was previously developed in this study. This includes a range of covariates. Figure 2 only includes the fetal activity. Why is the full model not included in the results?

3. There is a large overlap in results presented in the txt and in tables. It would improve the manuscript if the txt rather expands on than repeats the information in the table.

4. As the authors state in the "Interpretation" the type and quality of fetal movements change with advancing gestation. Did the authors consider stratified analyses for women (cases and controls) who were reporting fetal activity in late third trimester compared with (cases and controls) early third trimester, to see if the effect of strength and frequency differed from earlier in pregnancy?

REVIEWER	Valerie Smith
	Trinity College Dublin, Ireland
REVIEW RETURNED	15-Nov-2017

GENERAL COMMENTS	This is a well written, interesting and important paper that goes beyond outcomes associated with RFM and explores maternal perceptions and associated variables.
	In the main the methods are very well described, however, and even after reading the methods from the earlier paper (reference 13), I am not entirely clear as to how participants for the control arm were matched and approached; e.g. was the gestation for interview allocated at the time they were randomly recruited from the booking list or was this gestation determined by matching at the time the case presented? A line or two to clarify/explain this further would help; and how the researchers tracked this.
	Was there anything in the recruitment strategy that might have impacted on the low control uptake? Perhaps address how you attempted to minimise potential design recruitment limiters in this group; or any other thoughts on same.
	Line 206-207; states 'more likely' to have attended hospital, but the p-value of $p=0.07$ would not appear to support this; needs a check
	The Discussion is relevant and interesting, but it is quite lengthy and

could be shortened a bit, I feel.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
The manuscript is reporting fetal activity in pregnancies from women who have experienced stillbirth (cases) compared with gestational age matched women with ongoing pregnancies (controls). The choice of the case-control design is well justified by the authors and risk of recall bias is discussed in the paper. The research team is strong with long-term commitment to fetal movement research. I will commend the team for including additional characteristics of fetal movement such as strengths of the movement and hiccups and not only the frequency. This has been far less studied and may provide important input on how to inform women on when to seek medical attention.

However, I have concerns about the manuscript in its current form:
1. I am especially concerned about the combined frequency/strength variable that was constructed from Supplemental Table 1. This is not clear to me. As far as I understand it, Supplemental Table 1 was formed through a regression that essentially interacted categorical frequency with categorical strength. Similar odds ratios (e.g. 0.2-0.5, 2.0-3.0) were grouped together, and then the respective frequency/strength combinations were grouped together to form a more compact variable, which was then run in a secondary regression analysis (the main analysis for this paper).

Supplementary table 1 displays the 16 possible categories from the 4 levels each of the strength and frequency variables. Using this variable in this form would not be useful as the large number of categories would result in a lack of power. The compacted variable was created based on a priority of the prevalence of the most common scenario in the controls; this was done in conjunction with the odds ratio to ensure that variables with similar risks were combined. The supplementary table provides this information to the reader, so that they can interpret the primary data and the derivation of the variable. The reasons for generating this combined variable and a clearer description of how it was generated have been added to the methods section lines 130-136.

If correctly understood, I am not sure this is an appropriate analysis strategy. When you form this more compact variable, you already know the results of the regression. You have already "peeked behind the curtain", so to speak, and thus this new variable will be inherently biased.

We accept that the analysis could be seen as being exploratory, but not carrying out and publishing the results of such an analysis would lead to a situation where future researchers are unable to test/replicate them as they would be unknown. Whilst we understand the reviewers' point of view, they also point out that little work has been published in relation to fetal movements other than in relation to decreased fetal movements. There was little understanding prior to this study of the relationships of strength and frequency to each other. Epidemiological studies often require deeper investigation of the data in relation to results found to better understand them as is the case here. This is different to a RCT where the specified testing of groups is clearly essential. The authors do not believe that the new variable is biased, although it may overestimate the association, which needs to be tested in another dataset.

The hypothesis tests that have been performed are appropriate for testing a-priori hypotheses. To my understanding, Supplemental Table 1 is an a-priori hypothesis. The compact/combination variable is not an a-priori hypothesis, and as such, it is not appropriate to analyse this variable.
I see three available options:
a) Use supplemental Table 1 (or something similar) as your results (i.e. interact, but don't make it compact/grouped)
As stated above the study would not have adequate power to assess the effects of doing this due to the large number of degrees of freedom.
b) Run a regression without interacting frequency/strength (i.e. just run it as two variables that are each weaker/no change/stronger)

We considered this approach but feel that the understanding of the importance of the combinations of strength and frequency would be poorly understood by many clinicians using this approach.

c) Keep your compact/grouped variable, but validate it in a new independent cohort (in which this compact/grouped variable would be able to be tested in an a-priori manner without any bias issues)

We believe that this is the most appropriate approach and are working in collaboration with other cohorts to validate this finding. Other cohorts obviously would not wish to be subsumed into another study and thus we would expect that they would validate and publish separately. We have added this information to the discussion section of the manuscript (lines 264-267).

If I have understood the methods correctly, the manuscript needs to be re-analyzed before it is acceptable for publication.

We believe as stated above that the exploratory analyses should be presented to allow replication or otherwise in other cohorts.

2. It appears from the Methods that the fetal activity data were added to a model that was previously developed in this study. This includes a range of covariates. Figure 2 only includes the fetal activity. Why is the full model not included in the results?

The reviewer is correct that fetal activity was added to the model previously published (Heazell et al. BJOG. 2018 Jan;125(2):254-262). As noted in the manuscript this had little impact on the model and as this model has previously been published and the focus of the current paper is on fetal movement we don't believe it is necessary to effectively re publish those tables.

3. There is a large overlap in results presented in the txt and in tables. It would improve the manuscript if the txt rather expands on than repeats the information in the table.

The text has been shortened to avoid repetition of the results displayed in the table. The 95% confidence intervals have all been removed from the text. The results section is now 951 words (reduced from 1059). The statement comparing decreased fetal movements with increased fetal movements has been removed.

4. As the authors state in the "Interpretation" the type and quality of fetal movements change with advancing gestation. Did the authors consider stratified analyses for women (cases and controls) who were reporting fetal activity in late third trimester compared with (cases and controls) early third trimester, to see if the effect of strength and frequency differed from earlier in pregnancy?

Thank you for this suggestion. Unfortunately this study does not have the power to assess the effect of gestation. However, we are part of an individual pooled data consortium that has been developed and interactions such as this are planned to be assessed in that study. This has been addressed in the discussion section (lines 264-267).

Reviewer: 2

This is a well written, interesting and important paper that goes beyond outcomes associated with RFM and explores maternal perceptions and associated variables. In the main the methods are very well described, however, and even after reading the methods from the earlier paper (reference 13), I am not entirely clear as to how participants for the control arm were matched and approached; e.g. was the gestation for interview allocated at the time they were randomly recruited from the booking list or was this gestation determined by matching at the time the case presented? A line or two to clarify/explain this further would help; and how the researchers tracked this.

A more detailed explanation of how controls were approached has been included in the revised manuscript (lines 104-108). To clarify controls were not directly matched to cases. The distribution of the gestational age of controls was determined from the distribution of the gestational age of stillbirths in participating institutions in the preceding four years. Therefore, to ensure that the gestational age at interview was similar for cases and controls potential controls were identified from the booking list and randomly selected (to avoid selection bias) and the date of interview determined. This approach ensured that there was minimal difference between the average gestation of women in the control and case groups.

Was there anything in the recruitment strategy that might have impacted on the low control uptake? Perhaps address how you attempted to minimise potential design recruitment limiters in this group; or any other thoughts on same.

We have explored reasons for participation and non-participation in detail in a small sub-set of women this manuscript is currently under preparation. Some non-participants in the control group held negative views of research while others did not want to participate in a "stillbirth" study. Women were approached by the midwives providing clinical care which may have affected the amount of information they received about the study, or indeed the views of the staff may have had an impact. To address the reviewers' comment this issue has been covered in more detail in the discussion section (lines 245-250).

Line 206-207; states 'more likely' to have attended hospital, but the p-value of p=0.07 would not appear to support this; needs a check.

This text has been revised to say that women tended to have attended hospital more but this did not reach statistical significance (lines 205-206).

The Discussion is relevant and interesting, but it is quite lengthy and could be shortened a bit, I feel.

The discussion has been edited in accordance with the reviewers' suggestions. However, due to the recommendations of reviewers for additional components in the discussion section we have not been able to reduce the number of words in this section.

I confirm that this manuscript is not currently submitted for publication elsewhere and that none of the authors have any financial conflicts of interest to declare. We believe that this paper will be of significant interest to the international readership of BMJ Open and are timely given the Lancet Ending Preventable Stillbirth Series calling for efforts to reduce stillbirth in high-income countries and the NHS England Saving Babies Lives campaign which specifically addresses reduced fetal movements.

VERSION 2 – REVIEW

REVIEWER	Brita Askeland Winje Norwegian Institute of Public Health, P.O.Box 4404, Nydalen 0403 Oslo, Norway
REVIEW RETURNED	19-Jan-2018

GENERAL COMMENTS	Dear editor and authors Thank you for the opportunity to review a revised version of the manuscript entitled "Alterations in maternally-perceived fetal movement and their association with late stillbirth – findings from the Midland and North of England Stillbirth Case-Control Study". I thank the authors for the comprehensive response to my comments. However, I still have concerns related to the combined frequency/strength variable that was constructed from Supplemental Table 1. In my opinion, this has not been sufficiently addressed in the revised version. In the response letter the authors seem to agree with my concerns about the methodological approach, but argues that this should still be published so that other researchers may replicate the findings. If the aims of the author is to present exploratory analyses that can be replicated in other cohorts, then the authors should present supplemental Table 1 as their main analyses, and mention in the discussion about possible hypothetical groupings that could be validated in future studies. There is no need to perform the additional "prioritization groupings", because this may give false confidence intervals and significance results. If future authors wish to validate your hypothesized prioritization groups, they only need the information from Supplemental Table 1 to do so. If the authors want to further pursue option b, which was to run a

	regression without interacting frequency/strength, they can use the results from this regression and present post-hoc odds ratios for the strength/frequency groups that they think that clinicians would be interested in. With this, the regression would not contain any interaction terms, but the authors could manually calculate odds ratios and present them in a clinically interesting manner.
	Outside of these two options, I fail to see a statistically appropriate analysis and therefore disagree with the authors conclusion that the analysis should be published in its current form.
	I suggest to include a statistical review of the current manuscript

| **REVIEWER** | Valerie Smith
Trinity College Dublin, Ireland |
| **REVIEW RETURNED** | 31-Jan-2018 |

| **GENERAL COMMENTS** | Thank you for addressing my suggested amendments and resubmitting a revised manuscript. I am happy to recommend your paper for publication |

| **REVIEWER** | Javier Zamora
Hospital Ramón y Cajal, Spain |
| **REVIEW RETURNED** | 20-Mar-2018 |

GENERAL COMMENTS	I have reviewed statistical aspects of the manuscript only.
	Authors have used a standard statistical analysis to estimate the association of changes in either strength or frequency of fetal movements to the risk of late stillbirth. This is fine. However, It is worth noting that some decisions made during the analysis were heavily data-driven. First, the selection of variables for the multivariate model was driven by the results of univariate analyses. The association of these variables to the outcome were adjusted by a fixed list of covariates selected as a consequence of a previous model (i.e. maternal age, ethnicity, parity, education, smoking in pregnancy, marital status, birthweight centile and sleep factors). Second and more controversial is the creation of a composite exposure variable that is driven by the frequency and the association (OR) of the 16 potential combinations of strength and frequency of fetal movements. Supplementary table 1 is designed to illustrate how authors grouped the 16 levels of exposures. This grouping has been done post-hoc, i.e. after observing the magnitude of the associated risks. Doing so, we cannot disregard the presence of an important information bias that could well have biased results against the null.
	At this point, it is difficult to remove this bias. A sensitivity analysis would have provided an assessment of how critical this prioritization has been. Authors are encouraged to discuss this bias along with other limitations in the discussion section.

VERSION 2 – AUTHOR RESPONSE

Reviewer#1. If the editors consider that additional statistical review is required, we would be grateful if this could be obtained in a timely manner as it is now almost 6 months since our initial submission and over 18 weeks in peer review.

6

We note that reviewer 2 is satisfied with the revised manuscript. We also note that Reviewer 1 remains concerned about the development of a combined variable in the analyses and requested a further statistical review. With respect, we disagree with Reviewer 1 and agree with the statistical reviewer (Reviewer 3) who acknowledges that this was a data driven approach, and believes that we should address this as a limitation in the discussion rather than remove the supplementary table. We have added an additional section to this effect (lines 261-266).

As pointed out by Reviewer 1 we accept that the creation of this variable is an exploratory analysis, data driven and acknowledge that there are potentially inherent biases in doing this. However, exploratory analyses are often used (but probably not so obviously) in the categorisation of variables e.g. otherwise every paper would use the same categorisation of variables such as maternal age. We argue that such exploratory analysis can ensure the correct use of a variable. For example, the relationship of a continuous variable in relation to a binary outcome is often fitted in a model using a linear term and the true association missed if exploratory analysis to determine the true shape of the relationship is not carried out using something like a generalised additive model.

The authors stress that the current approach to resolving this issue is one of the options that Reviewer 1 suggested in their initial review (to keep and present the current variable and test it in a separate cohort). This replication has now been undertaken as part of the analysis of another study that will be published independently. Again this has been highlighted in the discussion section (lines 264-266).

VERSION 3 – REVIEW

REVIEWER	Javier Zamora Hospital Ramón y Cajal
REVIEW RETURNED	10-Apr-2018

GENERAL COMMENTS	I'm satisfied with the responses provided by the authors. They have acknowledged the limitations i suggested in my first review.

REVIEWER	Brita Askeland Winje Norwegian Institute of Public Health
REVIEW RETURNED	25-Apr-2018

GENERAL COMMENTS	Dear Dr Groves Thank you for the opportunity to review this revised version of the manuscript. I have previously expressed my concerns related to the data-driven analysis of the combined frequency/strength variable. This approach may lead to important information bias. In the revised version the author's have included a sentence in the discussion section to clarify the potential bias following this method. This is not mentioned in the Strengths and limitations in the introduction (bullet points). I will leave to the editor to decide upon publication.

VERSION 3 – AUTHOR RESPONSE

Reviewer#1 requests one further change to the manuscript with both Reviewer#2 and Reviewer#3 content with the content and presentation of the manuscript. We would be grateful if this revised

manuscript could be reviewed in a timely manner as it is now over 6 months since our initial submission and over 24 weeks in peer review.

We note that the editor requests that we improve the reporting of the statistics throughout the manuscript. As requested we have included the 95% CI for the reporting of all of the odds ratios in the text (changes are shown in blue font).

Reviewer#1 requests that we include the limitation of using a data driven variable in the Strengths and limitations in the introduction (bullet points). We have added this to this section and removed an earlier bullet point - Lines 62-64.

We hope that these changes allow this manuscript to be accepted for publication in