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)	Social, biological, behavioural and psychological factors related to physical activity during early pregnancy in the Screening for Pregnancy Endpoints (Cork, Ireland) cohort study
AUTHORS	Flannery, Caragh; Dahly, Darren; Byrne, Molly; Khashan, Ali; Mc Hugh, Sheena; Kenny, Louise; McAuliffe, Fionnuala; Kearney, Patricia

VERSION 1 - REVIEW

REVIEWER	Ruben Barakat Universidad Politécnica de Madrid. Spain
REVIEW RETURNED	03-Aug-2018

GENERAL COMMENTS	The authors present a cross-sectional study examining the different social, biological, behavioural and psychological factors that are linked with physical activity levels during pregnancy. The authors concluded that some key un-modifiable links that should guide the development of interventions, using a population approach, in order to encourage pregnant women to engage in Physical Activity during pregnancy. I found a study very carefully created and developed, the question posed by the authors is well defined: Page 4, lines 41-45 "Recognising and understanding the correlates of physical activity, as well explaining how these correlates influence subsequent behavior is fundamental to intervention development and implementation." The methods are appropriated. The discussion is well balanced and adequately supported by the data. However I have Minor Corrections for the authors. METHODS Page 6, line 8. Psychological and behavioural measures. Why did you use the Edinburgh Postnatal Depression Scale to assess depression during pregnancy and did not use Center for Epidemiological Studies Depression scale (CES-D)? RESULTS In page 7, line 47: "A total of 2579 nulliparous women were invited to participate in the SCOPE Irish study…." Please consider use of a Flow Chart of the Study participants. Page 9. Table 1: You present "Employment status" presenting only two categories (not working/working), however (in my opinion) this is an important limitation of the study. When you examining the occupational activity of the pregnant population the most recommendable categorization of the variable would have been: Housewives/Active work/Sedentary work. BMI category is corresponding to pre-pregnancy? Were not there any underweight pregnant women in the study?

REVIEWER	Fernanda G Surita
	University of Campinas, Brazil
REVIEW RETURNED	31-Aug-2018

GENERAL COMMENTS	Firstly, I would like to thank you for the opportunity to review an article submitted to the BMJ Open. The theme is interesting and timely in the area of women's health because I consider pregnancy an overload for the mother's body but also a window of opportunities for health education and positive actions.
	Title:
	Social, biological, behavioural and psychological factors and physical activity during pregnancy: a cross-sectional study
	The title should show the relation between the other variables and the physical activity, as it is placed in the objective of the summary - I suggest to review
	Abstract
	-higher socioeconomic status (≥24) - (≥24) ??? this does not make sense to me and many other readers
	-
	- PA subgroups identified: low PA (n = 393); moderate PA (n = 960); and high PA (n = 413) - there are well-established classifications for the degree / intensity of physical activity. If any of them were used it should appear from the abstract
	-
	Introduction
	A little long, the first two paragraphs do not bring anything new and could be summarized and mixed
	Method
	I was confused by the presentation of the physical activity levels of the abstract, but this question was adequately explained in the methods section
	Consumption of alcohol and smoking per day or per week or all pregnacie, please clarify
	Results
	Table 2 - remake
	Table 2 needs to be reviewed for reference groups.
	The reference group for comparison and risk assessment should be the most prevalent or least risk group.
	For example: 1- In age: Why are the reference group women under the age of 25 ?; 2- in ethnicity: Why are non-Caucasian women the reference women?
	It is necessary to review reference group 'for all variables
	The reference can not be a mathematical / statistical question and is considered the first group in the table, the reference in clinical practice has another sense, as quoted above.
	The socioeconomic condition of women needs to be better explained, as it only makes sense regionally, if an article to be read around the world this needs to be further explained.
	There are few notes that need to be revised and I believe they can be easily done by the authors in a new revised version
	Best regards

	*I can not access the statistical model used to calculate the levels of physical activity, so I include the need of the statistician. But this question can be reassessed and be at the discretion of the editor-in-chief

REVIEWER	Gordon Prescott
	University of Central Lancashire, UK
REVIEW RETURNED	21-Dec-2018

| GENERAL COMMENTS | I would like a little more focus on the direction of the effects in the abstract. I think this could be achieved by rephrasing a couple of lines in the results or conclusions. The conclusion in the main body of the paper is much stronger than the conclusion in the abstract.

It is possible to be confused by the reporting of the design and the linked strobe checklist. The data are clearly from a prospective cohort as reported in the text and prominently in the abstract. However, the introduction and methods (p5) describe it as cross sectional study as data from a single time point within pregnancy is used (apart from mode of delivery). The strobe checklist is completed following the options for a cross sectional study. So design presented in the abstract is inconsistent with the design in the strobe checklist. I think the authors need to resolve this and would favour retaining the original design rather than presenting this as a cross sectional study.

The description of the results from the multivariable logistic regression model is quite brief. The result for obesity does not mention that it is moderate vs. low (end of p10).

The choice to use the smaller category of two as the reference category is unusual, particularly in the case of non-Caucasian ethnicity, single marital status and smoking. There are very few non-Caucasian participants (n=41) and only 3 in the high activity group. This is reflected in the wide confidence intervals for ethnicity and makes me concerned about the stability of the models for high activity vs. low. The marked difference between the estimates for ethnicity in the high vs. low and moderate vs. low may reflect a genuine underlying difference in physical activity behaviour, but may be partly a consequence of small numbers. Caution is needed in the interpretation and recommendations about targeting interventions on the basis of ethnicity.

Many measured or ordinal variables have been re-categorised as binary which loses some of the detail in the collected data and some statistical power. This simplification should be acknowledged as an extra limitation or justified within the discussion of using potentially biased self-reported data.

Minor issues
Typos
P4 last line gained not gain.
P13 line 5 smokers not smoker.
The spacing around "n= ", "n =", "n=" or "n = " varies throughout text and tables. There is a similar issue around hyphens. |
|---|---|

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Ruben Barakat
Institution and Country: Universidad Politécnica de Madrid. Spain
Please state any competing interests or state 'None declared': None declared

Reviewer comments	Response to reviewer	Page & Line
The authors present a cross-sectional study examining the different social, biological, behavioural and psychological factors that are linked with physical activity levels during pregnancy. The authors concluded that some key un-modifiable links that should guide the development of interventions, using a population approach, in order to encourage pregnant women to engage in Physical Activity during pregnancy. I found a study very carefully created and developed, the question posed by the authors is well defined: Page 4, lines 41-45"Recognising and understanding the correlates of physical activity, as well explaining how these correlates influence subsequent behaviour is fundamental to intervention development and implementation." The methods are appropriated. The discussion is well balanced and adequately supported by the data.	We thank the reviewer for their comments.	
METHODS Page 6, line 8. Psychological and behavioural measures. Why did you use the Edinburgh Postnatal Depression Scale to assess depression during pregnancy and did not use Center for Epidemiological Studies Depression scale (CES-D)?	This was secondary analysis using data from the SCOPE study (www.scopestudy.net) in which recruitment commenced in 2004 and completed in 2011. We were therefore restricted to using the variables collected in this study. SCOPE collected the EPDS as, at the time the study launched, it was the only questionnaire to have recognised clinical cut-off values that relate to risk of depression. Furthermore, it has been extensively studied during pregnancy and post-natally (1, 2)	Page 6, Line 202
RESULTS In page 7, line 47: "A total of 2579 nulliparous women were invited to participate in the SCOPE Irish study…." Please consider use of a Flow Chart of the Study participants.	As requested, we have now included a basic flow diagram – see figure 2	Figure 2

• Page 9. Table 1: You present "Employment status" presenting only two categories (not working/working), however (in my opinion) this is an important limitation of the study. When you examining the occupational activity of the pregnant population the most recommendable categorization of the variable would have been: Housewives/Active work/Sedentary work.	We agree with the reviewer. The variable 'employment' was originally coded as (full time work; part time work; student; homemaker; unemployed; other). This was recoded to working vs. not working for our analysis. This is not an ideal combination but it wasn't possible to ascertain how active/sedentary employment was and due to small number in the categories we thought it best to code as a binary variable. We used this as a social indicator rather than activity. We have now acknowledged this in the limitations section of the paper.	Page 16, Line 396-400

Page 6, Line 183-185 |
| • BMI category is corresponding to pre-pregnancy?

• We're there any underweight pregnant women in the study? | BMI relates to BMI measured at the first SCOPE visit at 15 weeks gestation. This has now been highlighted in the paper.

There are 22 women who were underweight – we have highlighted this in Table 1. As there were so few women in the underweight category, for the multinomial logistic regression analysis, underweight and normal were collapsed together. This is now highlighted in Table 2. | Page 9, Line 284, Table 1; Page 12, Table 2 |

Reviewer: 2
Reviewer Name: Fernanda G Surita
Institution and Country: University of Campinas, Brazil
please state any competing interests or state 'None declared': None declared

Reviewer comments	Response to reviewer	Page & Line
Firstly, I would like to thank you for the opportunity to review an article submitted to the BMJ Open. The theme is interesting and timely in the area of women's health because I consider pregnancy an overload for the mother's body but also a window of opportunities for health education and positive actions.	We thank the reviewer for their comments and agree that the theme is interesting and timely.	
Title
Social, biological, behavioural and psychological factors and physical activity during pregnancy: a cross-sectional study		

• The title should show the relation between the other variables and the physical activity, as it is placed in the objective of the summary - I suggest to review	The title has been amended to reflect the relation between the other variables and physical activity. Furthermore, based on another reviewers comment the location of the study has been included in the title. 'Social, biological, behavioural and psychological factors related to physical activity during early pregnancy in the Screening for Pregnancy Endpoints (Cork, Ireland) cohort study'	Page 1, Line 1-2
Abstract • Higher socioeconomic status (≥24) - (≥24)??? this does not make sense to me and many other readers	Socioeconomic index (SEI) was based on the New Zealand SEI (<24 vs. ≥24) with higher values reflecting greater social status(3).The New Zealand SEI is an occupation-based measure of socio-economic status and it was used because SCOPE was an multicentre prospective cohort study in Auckland, New Zealand, Adelaide, Australia, Cork, Ireland, and Manchester, Leeds, and London, UK.	Page 2, Line 56-57; Page 6, Line 179-180
• PA subgroups identified: low PA (n = 393); moderate PA (n = 960); and high PA (n = 413) - there are well-established classifications for the degree / intensity of physical activity. If any of them were used it should appear from the abstract	No previously established classification for physical activity was used. Our classification was based on results from the latent class analysis. Based on a combination of model fit, parsimony, theoretical interpretability, and classification quality, the authors agreed that a three-class model was the most appropriate one. The three physical activity subgroups thus identified were: low levels of physical activity, moderate levels of physical activity and high levels of physical activity	Page 5-6, Line 157-171
Introduction • A little long, the first two paragraphs do not bring anything new and could be summarized and mixed	These two paragraphs have been condensed from 221 words to 160 and into one paragraph.	Page 4, Line 103-115
Method • I was confused by the presentation of the physical activity levels of the abstract, but this question was adequately explained in the methods section	We are glad that aspects of the methods sections clarified the physical activity levels.	

• Consumption of alcohol and smoking per day or per week or all pregnancies, please clarify •	Consumption of alcohol was measured as drinks per week and smoking as number of cigarettes per day at 15 weeks gestations at the first SCOPE visit– this has now been clarified in the methods section of the paper.	Page 6, Line 190-192
Results • Table 2 - remake • Table 2 needs to be reviewed for reference groups. • The reference group for comparison and risk assessment should be the most prevalent or least risk group.	We have changed the reference category for ethnicity (making Caucasians the reference group).	Page 12, Table 2
For example: 1- In age: Why are the reference group women under the age of 25?; 2- in ethnicity: Why are non-Caucasian women the reference women? It is necessary to review reference group 'for all variables. The reference cannot be a mathematical / statistical question and is considered the first group in the table, the reference in clinical practice has another sense, as quoted above.	We have now changed the reference category for ethnicity (making Caucasians the reference group). We can omit ethnicity due to lack of variability if needs be. While we agree with the reviewer's explanation of reference groups, we are trying to take a more positive approach by focusing on factors that improve chances of moderate or high PA. Furthermore, as this paper is framed in terms of factors that are associated with physical activity in pregnancy, the reference groups were selected based on this.	Table 2
• The socioeconomic condition of women needs to be better explained, as it only makes sense regionally, if an article to be read around the world this needs to be further explained. There are few notes that need to be revised and I believe they can be easily done by the authors in a new revised version I cannot access the statistical model used to calculate the levels of physical activity, so I include the need of the statistician. But this question can be reassessed and be at the discretion of the editor-in-chief	While the data reflects the socio economic condition of the women in Cork, Ireland, the measure used for the socioeconomic index (SEI) was based on the New Zealand SEI (<24 vs. ≥24) with higher values reflecting greater social status (3). New Zealand SEI is a validated scale and provides a link to international standards in socioeconomic status (4).	Page 6, Line 179-180

Reviewer: 3
Reviewer Name: Gordon Prescott
Institution and Country: University of Central Lancashire, UK
please state any competing interests or state 'None declared': None declared

Reviewer comments	Response to reviewer	Page & Line
Abstract - I would like a little more focus on the direction of the effects in the abstract. - I think this could be achieved by rephrasing a couple of lines in the results or conclusions. - The conclusion in the main body of the paper is much stronger than the conclusion in the abstract.	We agree with the reviewers comment and have included more detail in the abstract. We have now rephrased the results and conclusion.	Page 2, Line 52-60
- It is possible to be confused by the reporting of the design and the linked strobe checklist. The data are clearly from a prospective cohort as reported in the text and prominently in the abstract. However, the introduction and methods (p5) describe it as cross sectional study as data from a single time point within pregnancy is used (apart from mode of delivery). - The strobe checklist is completed following the options for a cross sectional study. So design presented in the abstract is inconsistent with the design in the strobe checklist. I think the authors need to resolve this and would favour retaining the original design rather than presenting this as a cross sectional study.	We agree with the reviewers comments and have removed all mention of a cross sectional study. This paper now sticks with the original design and its described as a secondary analysis of data from a prospective cohort study The strobe checklist used was the STROBE checklist for cohort, case-control, and cross-sectional studies (combined) and this has now been amended in supplementary material 1	Page 5, Line139-140; Line 152-155 Table S1
Results - The description of the results from the multivariable logistic regression model is quite brief. - The result for obesity does not mention that it is moderate vs. low (end of p10).	We thank the reviewer for this comment. We have now included more detail on the multivariable logistic regression model in the results section. We now discuss model 1 and 2 as well as the fully adjusted model. This has now be amended and reads - Of the biological factors, the relative risk for obese women (BMI >30kg/m^2) would be expected to increase (RRR 1.49 [95% CI: 1.00-2.22]) for moderate physical activity relative to normal (BMI <24 kg/m2) (vs. the low).	Page 10-11, Line 295-299 Page 10, Line 306-308

• The choice to use the smaller category of two as the reference category is unusual, particularly in the case of non-Caucasian ethnicity, single marital status and smoking. There are very few non-Caucasian participants (n=41) and only 3 in the high activity group. This is reflected in the wide confidence intervals for ethnicity and makes me concerned about the stability of the models for high activity vs. low. The marked difference between the estimates for ethnicity in the high vs. low and moderate vs. low may reflect a genuine underlying difference in physical activity behaviour, but may be partly a consequence of small numbers. • Caution is needed in the interpretation and recommendations about targeting interventions on the basis of ethnicity.	Based on the reviewers' comments we have changed the reference category for ethnicity (making Caucasians the reference group). We agree that ethnicity is problematic from a statistical point of view. We can omit ethnicity due to lack of variability if needs be. Furthermore, as this paper is framed in terms of factors that are associated with physical activity in pregnancy, the reference groups were selected based on this (taking a more positive approach by focusing on factors that improve chances of moderate or high PA). We have acknowledged this is the limitations section of the paper. We have removed any reference to recommendation for interventions targeting ethnicity in the discussion section.	Page 12, Table 2; Page 16, line 400-402 Discussion, page 14-16
• Many measured or ordinal variables have been re-categorised as binary which loses some of the detail in the collected data and some statistical power. This simplification should be acknowledged as an extra limitation or justified within the discussion of using potentially biased self-reported data.	We agree with the reviewer's comment and have now included this in the limitation section of the manuscript. We highlight that a number of variables were re-categorised to avoid categories with very small numbers in the analysis and for easy interpretation and presentation of results. However, by doing this, some information is lost, so caution must be used when interpreting the results as the statistical power to detect a relation between the variables and the physical activity outcome was reduced.	Page 16, Line 396-400
Minor issues • Typos P4 last line gained not gain. ✓ • P13 line 5 smokers not smoker. ✓ • The spacing around "n= ", "n =", "n=" or "n = " varies throughout text and tables. ✓ • There is a similar issue around hyphens. ✓	We thank the reviewer for highlighting these minor issues. These have now been amended in the manuscript.	

References

1. Stewart RC, Umar E, Tomenson B, Creed F. Validation of screening tools for antenatal depression in Malawi—A comparison of the Edinburgh Postnatal Depression Scale and Self Reporting Questionnaire. J Affect Disord. 2013;150(3):1041-7.

2. Austin MP, Hadzi-Pavlovic D, Saint K, Parker G. Antenatal screening for the prediction of postnatal depression: validation of a psychosocial Pregnancy Risk Questionnaire. Acta Psychiatr Scand. 2005;112(4):310-7.

3. Galbraith C, Jenkin G, Davis P, Coope P. New Zealand socio-economic index 1996. User's Guide Wellington: Statistics New Zealand. 2003.

4. Davis P, McLeod K, Ransom M, Ongley P, Pearce N, Howden-Chapman P. The New Zealand Socioeconomic Index: developing and validating an occupationally-derived indicator of socio-economic status. Aust N Z J Public Health. 1999;23(1):27-33.

VERSION 2 – REVIEW

REVIEWER	Ruben Barakat
	Universidad Politecnica de Madrid. Spain.
REVIEW RETURNED	19-Feb-2019

GENERAL COMMENTS	I appreciate the effort made by the authors to fulfill my corrections. Wonderful work. Congratulations!

REVIEWER	Fernanda Garanhani Surita
	University of Campinas, Brazil
REVIEW RETURNED	09-Mar-2019

| GENERAL COMMENTS | I did not find the authors' responses letter to the reviewers, so I would like a more specific evaluation of a statistician to evaluate the categorization of the level of physical activity. This variable is the core of the article and there can be no doubt as to its characterization, and I am not able to evaluate the proposed statistical model.

"The authors then met to discuss the results and a final number of latent classes were selected based on model fit statistics (using Akaike information criterion (AIC) and Bayesian information criterion (BIC)), parsimony, theoretical interpretability, and classification quality. Once the final model was chosen, participants were assigned to their most likely class (i.e. 171their modal assignment). Latent class models were estimated using MPlus version 8.0." |
| --- | --- |

REVIEWER	Gordon Prescott
	University of Central Lancashire, UK
REVIEW RETURNED	07-Mar-2019

GENERAL COMMENTS	The revision is much clearer and addresses the comments from the three reviewers. I would suggest some very minor changes, two for clarity and the rest typos. I do not think these would need to be checked by reviewers and instead could be approved by the editor.
	The first sentence of the conclusions in the abstract does not clear give the direction of effect. I would suggest adding the word "lower" or "low" before "physical activity" in this sentence.
	In line 150 in Study Design and Population, the "1" has been dropped from 15+/-1. In the Covariates section there are a couple of double spaces mid-sentence (lines 179 and 190).
	In the top two lines of Table 1, the "n=" spacing is still not consistent, with one use of n= with no spaces and one with "n= " and two of "n =".
	In line 304 it would be useful to state somewhere that this text is referring to model 3, e.g. "(Table 2, model 3)".
	In line 306 of page 10 Caucasians does not need an apostrophe.
	In lines 343 and 344, the sentence, "Women with a high education may ... " jars as the word "level" has also been included until this point.
	In lines 369 and 373 double spaces appear mid-sentence.
	The changes to the limitations section are very clear and answer my earlier concerns.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 1

Reviewer Name: Ruben Barakat

Institution and Country: Universidad Politecnica de Madrid. Spain.

Please state any competing interests or state 'None declared': None declared

Please leave your comments for the authors below I appreciate the effort made by the authors to fulfill my corrections.

Wonderful work.

Congratulations!

Reviewer: 3

Reviewer Name: Gordon Prescott

Institution and Country: University of Central Lancashire, UK Please state any competing interests or state 'None declared': None declared

Please leave your comments for the authors below

The revision is much clearer and addresses the comments from the three reviewers. I would suggest some very minor changes, two for clarity and the rest typos. I do not think these would need to be checked by reviewers and instead could be approved by the editor.

Item	Changes	Page, Line
The first sentence of the conclusions in the abstract does not clear give the direction of effect. I would suggest adding the word "lower" or "low" before "physical activity" in this sentence.	We thank the reviewer for this suggestion and have included 'low physical activity' in the last sentence of the abstract	Page 2, Line 60
In line 150 in Study Design and Population, the "1" has been dropped from 15+/-1. In the Covariates section there are a couple of double spaces mid-sentence (lines 179 and 190).	We have now added in the 15±1 and we have removed the double spaces	Page 5, Line 150
In the top two lines of Table 1, the "n=" spacing is still not consistent, with one use of n= with no spaces and one with "n= " and two of "n =".	We have removed the double spaces	Page 9, Table 1
In line 304 it would be useful to state somewhere that this text is referring to model 3, e.g. "(Table 2, model 3)".	This now reads as 'Results from the fully adjusted model (Table 2, model 3)'	Page 10, Line 304
In line 306 of page 10 Caucasians does not need an apostrophe.	The apostrophe has been removed	Page 10, Line 306
In lines 343 and 344, the sentence, "Women with a high education may ... " jars as the word "level" has also been included until this point. In lines 369 and 373 double spaces appear mid-sentence	This sentence has been reworded to 'Women with a low education level and those of a lower socio-economic class are less active and should be the focus of intervention efforts.'	Page 14, Line 345 Page 15, Line 369, 373
The changes to the limitations section are very clear and answer my earlier concerns.		

Reviewer: 2

Reviewer Name: Fernanda Garanhani Surita Institution and Country: University of Campinas, Brazil
please state any competing interests or state 'None declared': None declared

Please leave your comments for the authors below

I did not find the authors' responses letter to the reviewers, so I would like a more specific evaluation of a statistician to evaluate the categorization of the level of physical activity.

Item	Changes	
I did not find the authors' responses letter to the reviewers, so I would like a more specific evaluation of a statistician to evaluate the categorization of the level of physical activity.	We are happy to provide you with our original response to the reviewers letter	
This variable is the core of the article and there can be no doubt as to its characterization, and I am not able to evaluate the proposed statistical model. "The authors then met to discuss the results and a final number of latent classes were selected based on model fit statistics (using Akaike information criterion (AIC) and Bayesian information criterion (BIC)), parsimony, theoretical interpretability, and classification quality. Once the final model was chosen, participants were assigned to their most likely class (i.e. 171their modal assignment). Latent class models were estimated using MPlus version 8.0."	We agree with the reviewer and have now included a supplementary file which includes the results from the model fit statistics and the final three class mode.	Supplementary file 4