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)	Protective factors for child development at age two in the presence of poor maternal mental health: results from the All Our Babies (AOB) pregnancy cohort
AUTHORS	McDonald, Sheila; Kehler, Heather; Tough, Suzanne

VERSION 1 - REVIEW

REVIEWER	Maria Melchior INSERM, France
REVIEW RETURNED	08-May-2016

GENERAL COMMENTS	This paper aims to examine the role of parental temperament and interactions with their child as protective factors with regard to children's development if the mother has mental health problems. The research question is original and very important, but the methods used in the present study do not in my view make it possible to address it. Specifically: 1. Maternal mental health risk is measured over the lifecourse, implying that the exposed group includes mothers who have mental health problems in pregnancy of while the child is growing up and others who had mental health problems in the past. These are probably two different sources of risk - environmental and also genetic, and the role of genetic risk factors should be mentioned as a possible mechanism in the text. Given evidence on the importance of timing as well as chronicity of maternal mental health problems with regard to chidren's development, the exposure used in this study is probably too broad (van der Waerden et al, 2015). I wondered reading this manuscript whether the exposure studied here doesn't primarily measure the severity and chronicity of mental health problems among mothers, but this is not specified nor discussed. Given the availability of multiple measures of maternal mental health, the authors may consider studying trajectories rather than a cumulative indicator. 2. The protective factors identified include optimism, relation happiness and parenting self-efficacy - what do those measures mean in mothers who are depressed? Or rather, isn't that mothers who report these positive characteristics were depressed in the past rather than after the child was borrn? 3. Characteristics of study participants vs. non-participants and sample representativity should be described. 4. Are the different protective factors associated to one another? One may expect that, implying that they should be studied similtaneously rather than separately.

	5. What are the control variables included in the analyses? Obviously, at the very least, it is important to adjust for factors such as family socioeconomic and demographic characteristics.

REVIEWER	Dr. Kellie Thiessen University of Manitoba Canada
REVIEW RETURNED	09-May-2016

GENERAL COMMENTS	Overall this manuscript is well written. The content is organized that creates a very fluid read for the reviewer. The objective is clear, concepts are well defined and the statistical approach appears reasonable. The results yield information that informs health providers of what modifiable factors should be targeted to decrease developmental delays in early childhood in order to optimize life-long health for these children. I would only recommend that a more solid case of knowledge translation be given. Is there something specific in the context of this community that could be next steps.

REVIEWER	Elizabeth Prado University of California Davis, USA
REVIEW RETURNED	14-Jun-2016

GENERAL COMMENTS	Review of the Manuscript: Protective factors for child development at age two in the presence of poor maternal mental health: results from the All our Babies (AOB) pregnancy cohort The manuscript reports a prospective cohort study in Calgary, Canada in which pregnant women were enrolled at < 25 wk gestation and followed until 2 years post-partum. Among a sub-set of mothers who experienced mental health risk and who completed parent-report developmental screening questionnaires at 2 years post-partum, several factors were found to be protective of risk of developmental delay, including higher parenting self-efficacy, higher social support, higher optimism, more relationship happiness, and positive child sleep patterns. The study is well-written and clearly presented, with a clear discussion of implications of the findings for interventions likely to reduce risk of developmental delay among children of mothers with mental health risk. However, I would recommend a few modifications to the manuscript before publication. COMMENTS: 1) The exact number of mothers who experienced mental health risk and who comprise the primary set of participants for analysis is not reported in the abstract or in the manuscript. On p. 9, line 16, it is reported that 27% of 1341 mothers experienced mental health problems, from which it can be calculated ~362 participants constituted the primary analysis. This number should be reported in the abstract and methods. It is a bit misleading to report only the number 1596 in the abstract. 2) The ASQ is designed as a screener for developmental delay, thus the cut-off of 1 SD below the mean is probably conservative to catch as many children as possible who might be experiencing delay.

	Instead of reporting that the ASQ shows excellent validity (p. 5, line 42), it would be more informative to report the positive predictive value that has been found for this cut-off. This will inform the reader how likely it is that a child who screens positive actually experiences developmental delay. This also applies to the BITSEA. If the PPV is poor, it would be appropriate to discuss this in the limitations section. 3) Was a measure of nurturing and stimulation from the home environment completed, such as the HOME screener? This has been found to be strongly associated with child development in many studies, thus is a very important factor to consider as a covariate or potential protective factor or mediator of associations with maternal mental health. If this is not available, it would be appropriate to discuss this in the limitations section. 4) Page 7, line 11, it is unclear how the 3000 women enrolled was reduced to 2106 eligible for two year follow-up. With regard to attrition, it is reported in the strengths and limitations section that certain socio-demographic characteristics of the sample were not representative of the population, however, I do not see that reported in the results section. 5) I would suggest the following modifications to the analysis and results section to strengthen the overall argument of the paper. First, report the main effect of maternal mental health problems on risk of developmental delay. If significant, this shows that mothers with mental health problems are an important group to target for intervention. Second, compare the factors that are protective in mothers with mental health problems to the factors that are protective in mothers without mental health problems. If the same factors emerge as equally protective in both groups, that would suggest universal intervention would be beneficial. However, if the odds ratios are higher in the sub-group of mothers with mental health problems, this would support the argument for targeted intervention. 6) I do not see a completed STROBE checklist, though I may have missed it.

REVIEWER	Anna Sarkadi Dept of Women's and Children's Health, Uppsala University, Sweden
REVIEW RETURNED	30-Jun-2016

GENERAL COMMENTS	Thank you for the opportunity to review this paper. General comments This is an interesting paper examining potential protective factors on child outcomes in the presence of maternal mental health problems in a community sample of mothers in Alberta, Canada. Although the longitudinal design is well suited to identify protective factors, the authors use measures collected both before and concurrently with the outcome, which makes the models less clear and conclusions more difficult to draw. In addition, no population data is provided for comparison with the sample and thus we don't know to what extent the results are generalizable. Finally, the mental health problem variable is rather broadly defined, with 27% of mothers qualifying

	into this category. With such a prevalent risk factor, one is bound to find statistically significant associations when testing a large number of potential protective factors. Yet, the risk for type I error is not discussed. Specific comments Methods Main outcome. I am concerned by the use of language here. Using the term "developmental delay" is potentially problematic as it gives the impression of a diagnosed developmental delay, when in fact both the BITSEA and the ASQ are screening instruments for social and emotional behavioural and developmental problems. It would in fact be quite alarming if 14-18% of a child population had developmental delay. I would prefer the authors used a different term (e.g. developmental problems) that is more in accordance with the original "social-emotional/behavioral problems and delays in social-emotional competence" definition of e.g. the BITSEA. Main exposure. Maternal mental health risk is broadly defined. Although I very much support the life course perspective taken by the authors, the current definition risks to be over-inclusive, as also shown by the fact that 27% of mothers were classified as high maternal mental health risk. With such a prevalent risk factor, one is bound to find statistically significant associations when testing a large number of potential protective factors, risking a Type I error. While history of sexual or physical abuse and depression are well-established risk factors, the same is not true for "any" mental health problems or "any" abuse. The authors should give a better rationale for this definition or consider an alternative definition. It would also be interesting to just look at the depression outcome at 12 and 24 months and see if any protective factors can be identified and if these differ based on timing of the outcome. Independent variables. Some of these "independent" variables are actually not independent – they are known to be confounded by maternal educational level and child temperament, as well as child development. Therefore, the use of e.g. child screen time, childcare arrangement, child physical activity level, parental reading & imitation game playing, maternal working status and community resource use as independent variables is problematic. Please review the independent variables used and present a rationale as to their independence and only use the ones where such independence has been established empirically (e.g. lack of sign. correlation to main outcome and exposure variable) or where it is theoretically well motivated. Results Please provide state- and national data on the sample characteristics for comparison. A very common problem in studies is that we lose mothers who are lone parents, have a lower education and a first generation immigrant background. Although by no means specific to this study, it is very problematic that we build our knowledge base of child development on self-reports of essentially white, middle-class, married or cohabiting women from the majority population. Providing national statistics would allow for ascertaining generalizability. Conducting subgroup analysis would allow us to see if the general results in the study hold up for subgroups not well represented in studies.

	Models. When modelling for protective factors for age 2 years child development outcomes (Tables 3-5) I would strongly argue that measures collected at the same time NOT be included in the model. E.g. concurrently measured parenting self-efficacy can be both a consequence, as well as a cause of child development problems. It is not possible to discern this in cross-sectional data. Please indicate the R2 for each model. Although usually not huge, getting an idea of how much of the variance we are able to explain is useful in driving the field forward. If the variance in our outcomes is consistently better explained by things we haven't measured we have to rethink our approaches. It is not clear which of the independent (please also see my previous comment) variables have been tested in the models. Please state if a model is adjusted or unadjusted and please state all factors included in the adjusted models, even if they are not significant. E.g. maternal education should be included as a covariate given its wide impact, even if it is not independently contributing to the model at hand. As per my suggestion regarding the main exposure it would be interesting to see how these models would play out for depression alone. Discussion The paragraph on social support provided by parenting groups could be supplemented by Swedish data on the effects of parenting groups offered within the well child health system, where indeed increased social networks were the only significant outcome observed (1). 1. Fabian H, Rådestad I, Waldenström U. Childbirth and parenthood education classes in Sweden. Women's opinion and possible outcomes. Acta Obstet Gynecol Scand. 2005;84(5):436-43.

REVIEWER	Alison Teyhan University of Bristol, UK
REVIEW RETURNED	06-Jul-2016

GENERAL COMMENTS	This is an interesting paper, with a clear and important research question. It concerns the development of young children whose mothers have poor mental health, and identifies factors positively associated with development in these children. However, the approach taken and the results presented are not always clear. With revision and clarification (primarily to the methods and results sections), it could be suitable for publication. Introduction 1. It would be helpful if the authors defined the age that children start kindergarten in Canada, as this term has different meanings in different countries (Page 4, Line 9).

	Methods 2. Were the subsample who completed the 2-year follow-up questionnaire representative of the original cohort (Page 5, Lines 25/26)? 3. Is there a word missing in this sentence: 'an overall composite measure of child development was derived from ? and defined as....' (Page 5, Line 46) 4. The authors mention that the definition of risk in each domain (one SD below mean) is in accordance with the scoring instructions, but it's not clear if the overall composite measure (scoring at risk in at least two of five domains) is also a standardised measure, or one that the authors have developed themselves? (Page 5, Line 46) 5. How did the authors choose which protective factors to include? (Page 6, Line 15) 6. Given the large number of factors considered, there is the issue of multiple testing. Given this, the authors' decision to use relatively high p-value cut-offs of p<0.10 in the bivariate tests, and <0.05 in the multiple regressions, should be justified - and some mention needs to be made of the possibility of type 1 errors (perhaps in limitations section of discussion). 7. Why were the protective factor variables all dichotomised - surely a lot of detail lost by doing so? 8. The modelling strategy is a bit unclear. As I understand it, a block of variables is added - and then any variables within it which are p>0.05 are removed. The next block is then added - and then variables within it which are p>0.05 are removed. If the p-values for variables in earlier block(s) increase to over the threshold as a result of subsequent block(s) being added, are these variables in the earlier blocks then removed too? 9. Maternal mental health risk is not the 'main exposure' in the analysis (page 6, line 3). If I understand correctly, maternal mental health was used to select the

	sample eligible for inclusion in the analysis [i.e. only the mothers who had poor mental health were eligible for inclusion]. Therefore, although all of the children in the analysis were exposed to poor mental health, I think it is misleading to say this is the main 'exposure' in the analysis - as this would usually be taken to imply a variable in your model. 10. Related to point 9, the authors need to be much clearer on their analysis sample. In the abstract and participant follow-up section of the methods, the authors refer to a sample size of 1596. However, analysis was restricted to mothers who reported poor mental health (page 6, line 41). It should be explicit how many woman-child dyads were included in this main analysis - it is misleading to quote a figure of 1596, when the analyses were restricted to a sub-sample of that. The sample size analysed should be stated in the abstract and earlier on in the methods section. 11. A flow diagram of sample size could be helpful. Going from 2106 eligible for two year follow up, to 1596 completed, to 1341with complete mental health data, to 362 (27% of 1341) with poor mental health, to X number with complete data on other variables..... This level of information should be explicit to readers, at the moment it is not. Results 12. In Tables 1 and 2, the N for each variable should be given separately. The range of Ns given at present (e.g. '1303-1596' in Table 1) is not helpful to the reader, as it requires them to add up the numbers in each category of a variable to determine which variables have the most missing data. 13. The text on page 7, and Table 1, detail the percentage of mothers who had depression, and the percentage who had anxiety. But the proportion classified as having 'high maternal mental health risk' (and therefore eligible for inclusion in study) is not given until

	page 9. More detail on the women with ‘poor mental health’ should be given in the results - i.e. what were the characteristics of the women who were defined as having ‘poor mental health’ in terms of their mental health measures (as described on page 6, lines 8-11) and in terms of the variables in Table 1. At present Table 1 gives a description of all the woman in the follow-up sample, but this is not really very informative (although if compared to the baseline sample who did not complete the follow-up, as mentioned in point 2, that would be useful as a supplementary table). The reader needs to know about the woman in the actual analysis sample (i.e. those with ‘poor mental health’). 14. Similarly, Table 2 needs to show the characteristics of the children in the main analysis (i.e. those whose mothers have poor mental health). How do they compare to children whose mothers did not have poor mental health? 15. Tables 3-5 do not state sample size. 16. The authors mention ‘combinations’ of protective factors at several points in the article. Did they consider interactions between them? 17. Within the analysis sample, there would be differing severities and timings of maternal mental health difficulties. Was this considered in analysis? Discussion 18. Are the factors found to be protective of child development in this sample of mothers with poor mental health similar to what you would expect in a general population? I remain unclear on why the sample was restricted (and hence substantially reduced) to only mothers with ‘poor mental health’. Did the authors consider doing the analysis on the full sample, and investigating interactions between the protective factors and maternal mental health to see if associations

	differed between mothers in poor and not poor mental health? If no differences were found, then you would reach the same conclusions but with a much larger sample size and hence greater power. If differences were found then this would justify the approach taken in this paper - but it is not clear if the authors considered this. Strengths and limitations 19. It is misleading to say ‘data were analysed from 1596 mother-child dyads’ - this was a study restricted to the children of mothers who had poor mental health. The small sample size should be mentioned as a limitation.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

Reviewer Name

Maria Melchior

Institution and Country

INSERM, France

Please state any competing interests or state ‘None declared’:
None

Please leave your comments for the authors below This paper aims to examine the role of parental temperament and interactions with their child as protective factors with regard to children's development if the mother has mental health problems. The research question is original and very important, but the methods used in the present study do not in my view make it possible to address it.

Specifically:

1. Maternal mental health risk is measured over the lifecourse, implying that the exposed group includes mothers who have mental health problems in pregnancy of while the child is growing up and others who had mental health problems in the past. These are probably two different sources of risk - environmental and also genetic, and the role of genetic risk factors should be mentioned as a possible mechanism in the text. Given evidence on the importance of timing as well as chronicity of maternal mental health problems with regard to chidren's development, the exposure used in this study is probably too broad (van der Waerden et al, 2015). I wondered reading this manuscript whether the exposure studied here doesn't primarily measure the severity and chronicity of mental health problems among mothers, but this is not specified nor discussed.
Given the availability of multiple measures of maternal mental health, the authors may consider studying trajectories rather than a cumulative indicator.

Response: Thank-you for this comment. We agree that the definition of this group is broad yet it aligns with previous research, now referenced in the text. We further discuss the notion of trajectories as this is the focus of further analyses in this cohort using more waves of data that were not the focus of the current study.

2. The protective factors identified include optimism, relation happiness and parenting self-efficacy - what do those measures mean in mothers who are depressed? Or rather, isn't that mothers who report these positive characteristics were depressed in the past rather than after the child was borrn?

Response: The protective factors that we examined in this analysis were drawn from different data waves representing the prenatal and postpartum period, and one could argue that they are intrinsic characteristics (e.g., optimism). We acknowledge in the Discussion that due to the self-report nature of these variables, they may indeed have a different meaning among women identified as having mental health risk.

3. Characteristics of study participants vs. non-participants and sample representativity should be described.

Response: Thank-you for this comment. We agree and have added this description to the text. In brief, attrition and other design limitations at follow-up waves excluded participants with lower SES and those not born in Canada.

4. Are the different protective factors associated to one another? One may expect that, implying that they should be studied similtaneously rather than separately.

Response: We agree that the protective factors are overlapping. Our multivariable regression analysis identified those that were independent protective factors.

5. What are the control variables included in the analyses? Obviously, at the very least, it is important to adjust for factors such as family socioeconomic and demographic characteristics.

Response: Our final models contain only those factors that remained significant in the model. Socioeconomic factors and demographic factors were examined but were not significant in the final model, nor did they confound the associations between the remaining independent variables and the outcome. Our modeling approach was to build a final, parsimonious model. We note in the test that our final models were not confounded by SES or other demographic variables.

Reviewer: 2

Reviewer Name

Dr. Kellie Thiessen

Institution and Country

University of Manitoba
Canada

Please state any competing interests or state ‘None declared’:
None declared

Please leave your comments for the authors below Overall this manuscript is well written. The content is organized that creates a very fluid read for the reviewer. The objective is clear, concepts are well defined and the statistical approach appears reasonable. The results yield information that informs health providers of what modifiable factors should be targeted to decrease developmental delays in early childhood in order to optimize life-long health for these children.
I would only recommend that a more solid case of knowledge translation be given. Is there something specific in the context of this community that could be next steps.

Response: Thank-you for this comment. We elaborate on specific potential strategies in the Conclusion.

Reviewer: 3

Reviewer Name

Elizabeth Prado

Institution and Country

University of California Davis, USA

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

Please leave your comments for the authors below Review of the Manuscript: Protective factors for child development at age two in the presence of poor maternal mental health: results from the All our Babies (AOB) pregnancy cohort

The manuscript reports a prospective cohort study in Calgary, Canada in which pregnant women were enrolled at < 25 wk gestation and followed until 2 years post-partum. Among a sub-set of mothers who experienced mental health risk and who completed parent-report developmental screening questionnaires at 2 years post-partum, several factors were found to be protective of risk of developmental delay, including higher parenting self-efficacy, higher social support, higher optimism, more relationship happiness, and positive child sleep patterns. The study is well-written and clearly presented, with a clear discussion of implications of the findings for interventions likely to reduce risk of developmental delay among children of mothers with mental health risk. However, I would recommend a few modifications to the manuscript before publication.

COMMENTS:
1) The exact number of mothers who experienced mental health risk and who comprise the primary set of participants for analysis is not reported in the abstract or in the manuscript. On p. 9, line 16, it is reported that 27% of 1341 mothers experienced mental health problems, from which it can be calculated ~362 participants constituted the primary analysis. This number should be reported in the abstract and methods. It is a bit misleading to report only the number 1596 in the abstract.

Response: Thank-you for this comment. We agree and report this frequency in the abstract and highlight that this is a subgroup throughout.

2) The ASQ is designed as a screener for developmental delay, thus the cut-off of 1 SD below the mean is probably conservative to catch as many children as possible who might be experiencing delay. Instead of reporting that the ASQ shows excellent validity (p. 5, line 42), it would be more

informative to report the positive predictive value that has been found for this cut-off. This will inform the reader how likely it is that a child who screens positive actually experiences developmental delay. This also applies to the BITSEA. If the PPV is poor, it would be appropriate to discuss this in the limitations section.

Response: Thank-you for this comment. Given that our cut-off of 1SD defined children 'in at least the monitoring zone' for delay, we cannot directly link this to a PPV (note: validity indices, including PPV has been reported for the 2 SD cut-off; we would expect the PPV for the 1SD cut-off to be lower). Given that the PPVs for these screeners are low, we acknowledge this as a limitation in the Discussion as suggested.

3) Was a measure of nurturing and stimulation from the home environment completed, such as the HOME screener? This has been found to be strongly associated with child development in many studies, thus is a very important factor to consider as a covariate or potential protective factor or mediator of associations with maternal mental health. If this is not available, it would be appropriate to discuss this in the limitations section.

Response: We agree with this limitation and have noted as such in the text. We did not use the HOME screener or a similar tool in our study.

4) Page 7, line 11, it is unclear how the 3000 women enrolled was reduced to 2106 eligible for two year follow-up. With regard to attrition, it is reported in the strengths and limitations section that certain socio-demographic characteristics of the sample were not representative of the population, however, I do not see that reported in the results section.

Response: Thank-you for this comment as it is being raised across all reviewers. We have addressed this in the Methods and Discussion sections.

5) I would suggest the following modifications to the analysis and results section to strengthen the overall argument of the paper. First, report the main effect of maternal mental health problems on risk of developmental delay. If significant, this shows that mothers with mental health problems are an important group to target for intervention. Second, compare the factors that are protective in mothers with mental health problems to the factors that are protective in mothers without mental health problems. If the same factors emerge as equally protective in both groups, that would suggest universal intervention would be beneficial. However, if the odds ratios are higher in the sub-group of mothers with mental health problems, this would support the argument for targeted intervention.

Response: Thank-you for this comment. The main effect analyses are the focus of a manuscript under review. In terms of protective factors among mothers WITHOUT mental health risk compared to women WITHOUT a mental health risk, we discuss these further analyses in the Discussion section. Indeed, there was evidence for suggesting targeted interventions, given that some factors were unique to each context.

6) I do not see a completed STROBE checklist, though I may have missed it.

Response: A STROBE checklist was completed and has been uploaded.

Reviewer: 4

Reviewer Name

Anna Sarkadi

Institution and Country

Dept of Women's and Children's Health, Uppsala University, Sweden

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

Please leave your comments for the authors below BMJ Open-2016-012096

Thank you for the opportunity to review this paper.

General comments
This is an interesting paper examining potential protective factors on child outcomes in the presence of maternal mental health problems in a community sample of mothers in Alberta, Canada. Although the longitudinal design is well suited to identify protective factors, the authors use measures collected both before and concurrently with the outcome, which makes the models less clear and conclusions more difficult to draw. In addition, no population data is provided for comparison with the sample and thus we don't know to what extent the results are generalizable. Finally, the mental health problem variable is rather broadly defined, with 27% of mothers qualifying into this category. With such a prevalent risk factor, one is bound to find statistically significant associations when testing a large number of potential protective factors. Yet, the risk for type I error is not discussed.

Response: Thank-you for these comments. We have addressed the issues of generalizability and attrition in the manuscript as they are common comments from all reviewers. As for the 27% subgroup, given the smaller sample size (n=305), we have low power to determine associations within this group, compared to the full sample. We feel that our multivariable model building process, including checks for robustness (now described), address concerns about analytic rigor.

Specific comments
Methods
Main outcome. I am concerned by the use of language here. Using the term "developmental delay" is potentially problematic as it gives the impression of a diagnosed developmental delay, when in fact both the BITSEA and the ASQ are screening instruments for social and emotional behavioural and developmental problems. It would in fact be quite alarming if 14-18% of a child population had developmental delay. I would prefer the authors used a different term (e.g. developmental problems) that is more in accordance with the original "social-emotional/behavioral problems and delays in social-emotional competence" definition of e.g. the BITSEA.

Response: We have changed developmental delay to developmental problems throughout as suggested.

Main exposure. Maternal mental health risk is broadly defined. Although I very much support the life course perspective taken by the authors, the current definition risks to be over-inclusive, as also shown by the fact that 27% of mothers were classified as high maternal mental health risk. With such a prevalent risk factor, one is bound to find statistically significant associations when testing a large number of potential protective factors, risking a Type I error. While history of sexual or physical abuse and depression are well-established risk factors, the same is not true for "any" mental health problems or "any" abuse. The authors should give a better rationale for this definition or consider an alternative definition. It would also be interesting to just look at the depression outcome at 12 and 24 months and see if any protective factors can be identified and if these differ based on timing of the outcome.

Independent variables. Some of these “independent” variables are actually not independent – they are known to be confounded by maternal educational level and child temperament, as well as child development. Therefore, the use of e.g. child screen time, childcare arrangement, child physical activity level, parental reading & imitation game playing, maternal working status and community resource use as independent variables is problematic. Please review the independent variables used and present a rationale as to their independence and only use the ones where such independence has been established empirically (e.g. lack of sign. correlation to main outcome and exposure variable) or where it is theoretically well motivated.

Response. Thank-you for this comment. We hope that our approach to multivariable model building addresses the concern of independence, and acknowledge the potential for misclassification and residual confounding of factors not measured or included, as suggested by the reviewer. We also acknowledge the risk for type 1 error given the large number of protective factors examined in the bivariate analysis.

Results
Please provide state- and national data on the sample characteristics for comparison. A very common problem in studies is that we lose mothers who are lone parents, have a lower education and a first generation immigrant background. Although by no means specific to this study, it is very problematic that we build our knowledge base of child development on self-reports of essentially white, middle-class, married or cohabiting women from the majority population.

Providing national statistics would allow for ascertaining generalizability. Conducting subgroup analysis would allow us to see if the general results in the study hold up for subgroups not well represented in studies.

Response: Thank-you for this comment. We address the concerns around generalizability and attrition as noted above. Differing sample characteristics among continuers and those recruited limit generalizability and generate a potential selection bias. These limitations are addressed in the text.

Models. When modelling for protective factors for age 2 years child development outcomes (Tables 3-5) I would strongly argue that measures collected at the same time NOT be included in the model. E.g. concurrently measured parenting self-efficacy can be both a consequence, as well as a cause of child development problems. It is not possible to discern this in cross-sectional data.

Response: We agree with this limitation and potential for reverse causality bias and acknowledge this in the text.

Please indicate the R2 for each model. Although usually not huge, getting an idea of how much of the variance we are able to explain is useful in driving the field forward. If the variance in our outcomes is consistently better explained by things we haven’t measured we have to rethink our approaches.

Response: Thank-you for this comment. Given that we used logistic regression we do not report an R2. We note the potential for residual confounding as noted above.

It is not clear which of the independent (please also see my previous comment) variables have been tested in the models. Please state if a model is adjusted or unadjusted and please state all factors included in the adjusted models, even if they are not significant. E.g. maternal education should be included as a covariate given its wide impact, even if it is not independently contributing to the model at hand.

As per my suggestion regarding the main exposure it would be interesting to see how these models would play out for depression alone.

Response: Thank-you for this comment. We highlight that final models are adjusted models, and we checked for robustness by examining the extent to which SES and demographics confounded associations. Only final, parsimonious models are presented. The examination of depression as a main effect is the focus of another manuscript under review.

Discussion

The paragraph on social support provided by parenting groups could be supplemented by Swedish data on the effects of parenting groups offered within the well child health system, where indeed increased social networks were the only significant outcome observed (1).

1. Fabian H, Rådestad I, Waldenström U. Childbirth and parenthood education classes in Sweden. Women's opinion and possible outcomes. Acta Obstet Gynecol Scand. 2005;84(5):436-43.

Response: Thank-you for this suggestion. We have added this reference in the Discussion.

Reviewer: 5

Reviewer Name

Alison Teyhan

Institution and Country

University of Bristol, UK

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

This is an interesting paper, with a clear and important research question. It concerns the development of young children whose mothers have poor mental health, and identifies factors positively associated with development in these children. However, the approach taken and the results presented are not always clear. With revision and clarification (primarily to the methods and results sections), it could be suitable for publication.

Introduction

1. It would be helpful if the authors defined the age that children start kindergarten in Canada,

as this term has different meanings in different countries (Page 4, Line 9).

Response: Thank-you for this comment. We have inserted that kindergarten in Canada refers to an approximate age of 5 years.

Methods

2. Were the subsample who completed the 2-year follow-up questionnaire representative of the

original cohort (Page 5, Lines 25/26)?

Response: Thank-you for this comment. We have addressed this important issue throughout as this was a comment raised by other reviewers.

3. Is there a word missing in this sentence: 'an overall composite measure of child development was derived from ? and defined as....' (Page 5, Line 46)

Response: Thank-you, this has been fixed.

4. The authors mention that the definition of risk in each domain (one SD below mean) is in accordance with the scoring instructions, but it's not clear if the overall composite measure (scoring at risk in at least two of five domains) is also a standardised measure, or one that the authors have developed themselves? (Page 5, Line 46)

Response: Thank-you for this comment. We have provided justification on our composite measure definition, namely that this classification aligns with an overall proportion of delay as seen in the literature, ranging from 12-16%.

5. How did the authors choose which protective factors to include? (Page 6, Line 15)

6. Given the large number of factors considered, there is the issue of multiple testing. Given this, the authors' decision to use relatively high p-value cut-offs of p<0.10 in the bivariate tests, and <0.05 in the multiple regressions, should be justified - and some mention needs to be made of the possibility of type 1 errors (perhaps in limitations section of discussion).

Response: Thank-you for this comment. We have acknowledged the potential for type 1 error in our bivariate analysis given the large number of protective factors examined.

7. Why were the protective factor variables all dichotomised - surely a lot of detail lost by doing so?

Response: We acknowledge this as a limitation in the Discussion.

8. The modelling strategy is a bit unclear. As I understand it, a block of variables is added - and then any variables within it which are p>0.05 are removed. The next block is then added - and then variables within it which are p>0.05 are removed. If the p-values for variables in earlier block(s) increase to over the threshold as a result of subsequent block(s) being added, are these variables in the earlier blocks then removed too?

Response: Thank-you for this comment. We have clarified our approach in the Methods section.

9. Maternal mental health risk is not the 'main exposure' in the analysis (page 6, line 3). If I understand correctly, maternal mental health was used to select the sample eligible for inclusion in the analysis [i.e. only the mothers who had poor mental health were eligible for inclusion]. Therefore, although all of the children in the analysis were exposed to poor mental health, I think it is misleading to say this is the main 'exposure' in the analysis - as this would usually be taken to imply a variable in your model.

Response: we completely agree and no longer refer to maternal mental health risk as an exposure variable.

10. Related to point 9, the authors need to be much clearer on their analysis sample. In the abstract and participant follow-up section of the methods, the authors refer to a sample size of 1596. However, analysis was restricted to mothers who reported poor mental health (page 6, line 41). It should be explicit how many woman-child dyads were included in this main analysis - it is misleading to quote a figure of 1596, when the analyses were restricted to a sub-sample of that. The sample size analysed should be stated in the abstract and earlier on in the methods section.
Response: We agree and have modified accordingly. We present the sample size for both the descriptive analysis and the resilience analysis upfront, and clarify that we were also interested in describing child development at age 2, followed by an examination of protective factors among a subgroup at higher risk for poor child outcomes.

11. A flow diagram of sample size could be helpful. Going from 2106 eligible for two year follow up, to 1596 completed, to 1341with complete mental health data, to 362 (27% of 1341) with poor mental health, to X number with complete data on other variables..... This level of information should be explicit to readers, at the moment it is not.

Response: Thank-you for this comment. We have attempted to clarify this issue both upfront and throughout the manuscript.

Results

12. In Tables 1 and 2, the N for each variable should be given separately. The range of Ns given at present (e.g. '1303-1596' in Table 1) is not helpful to the reader, as it requires them to add up the numbers in each category of a variable to determine which variables have the most missing data.

13. The text on page 7, and Table 1, detail the percentage of mothers who had depression, and the percentage who had anxiety. But the proportion classified as having 'high maternal mental health risk' (and therefore eligible for inclusion in study) is not given until page 9. More detail on the women with 'poor mental health' should be given in the results - i.e. what were the characteristics of the women who were defined as having 'poor mental health' in terms of their mental health measures (as described on page 6, lines 8-11) and in terms of the variables in Table 1. At present Table 1 gives a description of all the woman in the follow-up sample, but this is not really very informative (although if compared to the baseline sample who did not complete the follow-up, as mentioned in point 2, that would be useful as a supplementary table). The reader needs to know about the woman in the actual analysis sample (i.e. those with 'poor mental health').

Response: Thank-you for this comment. We have provided more detail on this subgroup in terms of the separate variables that define mental health risk and have clarified the sample size for each variable in the table as suggested.

14. Similarly, Table 2 needs to show the characteristics of the children in the main analysis (i.e. those whose mothers have poor mental health). How do they compare to children whose mothers did not have poor mental health?

Response: Thank-you for this comment. We have clarified that our interest was in both describing child outcomes at age 2 and examining protective factors among a subgroup of families deemed higher risk for poor child outcomes. Further we have provided the proportion of delay on all outcomes within the subgroup in the text.

15. Tables 3-5 do not state sample size.

Response: We have added the sample size in the tables.

16. The authors mention 'combinations' of protective factors at several points in the article. Did they consider interactions between them?

Response: Thank-you for this comment. We have removed the term 'combinations' to avoid confusion.

17. Within the analysis sample, there would be differing severities and timings of maternal mental health difficulties. Was this considered in analysis?

Response: Thank-you for this comment. We have addressed this in the Conclusion as this is the focus of another manuscript.

Discussion

18. Are the factors found to be protective of child development in this sample of mothers with poor mental health similar to what you would expect in a general population? I remain unclear on why the sample was restricted (and hence substantially reduced) to only mothers with 'poor mental health'. Did the authors consider doing the analysis on the full sample, and investigating interactions between the protective factors and maternal mental health to see if associations differed between mothers in poor and not poor mental health? If no differences were found, then you would reach the same conclusions but with a much larger sample size and hence greater power. If differences were found then this would justify the approach taken in this paper - but it is not clear if the authors considered this.

Response: Thank-you for this comment. We now present in the text results on protective factors among those WITHOUT maternal mental health risk and implications for targeted and universal strategies. We did not consider interactions given that our maternal mental health risk subgroup was a composite variable and defined for the purposes of this study. We appreciate this alternative approach and will consider in future work.

Strengths and limitations

19. It is misleading to say 'data were analysed from 1596 mother-child dyads' - this was a study restricted to the children of mothers who had poor mental health. The small sample size should be mentioned as a limitation.

Response: We agree and have been explicit in the manuscript that the sample size for the resilience analysis was a subgroup of the bigger sample at age 2.

VERSION 2 – REVIEW

REVIEWER	Elizabeth Prado University of California Davis, USA
REVIEW RETURNED	25-Aug-2016

GENERAL COMMENTS	The authors have addressed all comments adequately.

REVIEWER	Anna Sarkadi Dept of Public Health and Caring Sciences, Uppsala University
REVIEW RETURNED	17-Aug-2016

GENERAL COMMENTS	Thank you for the revised version of this article. The authors have taken most reviewer suggestions into account or provided

	explanations for when they did not. A notable exception is the inclusion of measures collected at the same time as predictive variables in the model. Thus, I still lack the rationale to include parenting self-efficacy as an independent variable in the model. The authors draw conclusions about this factor being so important as to warrant targeted interventions. However, most interventions that turn out to affect parenting self-efficacy aim at improving parenting practices and thus have an entry point of parents identifying behavioural or emotional problems in their children or general problems in their relationship with their child and not low self-efficacy. Thus, I believe it is not only inappropriate to include parenting self-efficacy measured concurrently into the model as a predictive independent variable – it also risks misguiding conclusions about appropriate interventions. Please also change the term "developmental delay" into developmental vulnerability in §1, 4th row when you cite the Hertzman paper. It is vulnerability and not delay the EDI describes.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 4
Please leave your comments for the authors below
1. Thank you for the revised version of this article. The authors have taken most reviewer suggestions into account or provided explanations for when they did not. A notable exception is the inclusion of measures collected at the same time as predictive variables in the model. Thus, I still lack the rationale to include parenting self-efficacy as an independent variable in the model. The authors draw conclusions about this factor being so important as to warrant targeted interventions. However, most interventions that turn out to affect parenting self-efficacy aim at improving parenting practices and thus have an entry point of parents identifying behavioural or emotional problems in their children or general problems in their relationship with their child and not low self-efficacy. Thus, I believe it is not only inappropriate to include parenting self-efficacy measured concurrently into the model as a predictive independent variable – it also risks misguiding conclusions about appropriate interventions.

Response: Thank-you for this comment. The research team has discussed this comment and has removed parenting self-efficacy from the analysis due to the considerations noted above and in order to avoid further misinterpretation by the reader. Specifically, we did not consider parenting self-efficacy an eligible protective factor for the outcome at age 2 due to the timing of its assessment with the outcome and potential for reverse-causality bias. We –re-ran the multivariable models in the resilience analysis. Predictive factors among mothers at high maternal mental health risk for the developmental vulnerability outcome changed from higher social support during pregnancy, more relationship happiness at one year postpartum and higher parenting self-efficacy at two years postpartum TO higher social support during pregnancy, and more relationship happiness at one year postpartum. Predictive factors among mothers at high maternal mental health risk for the social-emotional outcome changed from more relationship happiness at one year postpartum, higher parenting self-efficacy at two years postpartum and the child being able to fall asleep in less than 30 minutes at night TO more relationship happiness at one year postpartum, less difficulty balancing family, work and other responsibilities and the child being exposed to less than one hour of screen time per day. There were no changes for the third outcome, behavioral problems. These changes have implications for our Discussion and as such we recommend interventions that focus on these factors, and not parenting self-efficacy. Changes are highlighted in the marked copy of the submitted manuscript, including changes made to the Discussion section and Abstract.

2. Please also change the term “developmental delay” into developmental vulnerability in §1, 4th row when you cite the Hertzman paper. It is vulnerability and not delay the EDI describes.

Response: Thank-you for this comment. We agree with the reviewer and have made this change.

Reviewer: 3
1. The authors have addressed all comments adequately.

Response: thank-you for this comment.

We feel that the manuscript is more focused and decreases any risk for misinterpretation. We look forward to this contribution to the literature and thank the reviewers and editorial team for their insightful feedback.

VERSION 3 – REVIEW

REVIEWER	Anna Sarkadi Dept. of Public Health and Caring Sciences, Uppsala University, Sweden
REVIEW RETURNED	10-Oct-2016

GENERAL COMMENTS	Thank you for this final version of the paper. I think the paper is clear and recommendations well founded. Although I did not feel including self-efficacy in the model was justified I think it would be ok to present existing correlations (if any) between the variables finally included and self-efficacy as a way of showing that these factors are interrelated. Parenting programmes usually have good effect on child behavior problems, but also maternal mental health and self-efficacy as a "side-effect". In addition, strategies taught in parenting programs can help parents implement things, such as limited screen time and to find time to focus on the couple relationship. Finally, parenting groups can increase social networks as well. Therefore, I think recommending evidence-based parenting programs for this vulnerable population of mothers with mental health problems would be justified. This, however is just a suggestion, at the discretion of the authors to include.