

## 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

|                            |  |
|----------------------------|--|
| <b>TITLE (PROVISIONAL)</b> | Breastfeeding and behavioural problems: Propensity score matching with a national cohort of infants in Chile |
| <b>AUTHORS</b>             | Lisa-Christine , Girard; Farkas, Chamarrita  |

### VERSION 1 – REVIEW

|                        |   |
|------------------------|---|
| <b>REVIEWER</b>        | CHRISTIAN LORET DE MOLA<br>Federal University of Rio Grande, Brazil |
| <b>REVIEW RETURNED</b> | 06-Aug-2018   |

|                         |   |
|-------------------------|---|
| <b>GENERAL COMMENTS</b> | <p>First, I would like to congratulate the authors for the selected research subject. It is of interest, and an important addition to the literature and what it is known of the long term effects of breastfeeding. I would only like to make some general comments and some specific ones.</p> <p>General and specific comments</p> <p>Introduction</p> <p>The introduction could be a little more straight forward as well as the objective presented, you could explain your methodology in the methods section, I think you do not need to make a preview of it in the objective, and maybe discuss those points in the discussion section, and discuss the advantages, disadvantages and limitations of this method.</p> <p>Methods</p> <p>First paragraph of the methods section is clear until you start talking about the 4300 children who were included in the study, you said it were all who had data on the confounding, is this still in wave one? It was clear after 3-4 times of reading it that actually it was the same wave, and from the 15000 only this 4thousand had all the need data. Maybe rephrase a little to make it clearer. It is unclear until this point if this a longitudinal or cross sectional study, maybe you should start the paragraph stating this.</p> <p>Along the methods section, you start giving some explanations om why you used this method, or what literature says about the confounder, or ethical issues, this should be discuss in the discussion section, in the methods you should only describe what you did.</p> <p>You do not mention the means correlation you made with the CBCL scores. In addition, in my experience the CBCL normally is not normally distributed, therefore a mean comparison of correlation is not adequate. Same as the for the regression model used, comparing means is not adequate for the CBCL unless you demonstrated that the models complied to the basic requirements of a regression model.</p> <p>Results</p> <p>Results section should be re-written, it is not informative. Talking about being or not significant is not as relevant as telling us the</p> |
|-------------------------|---|

|  |  |
|--|--|
|  | <p>amount of difference and correctly showing this Betas. It is also lacking fundamental parts, like a description of our population in terms of the variables you used. Also interpretations and opinions should not be included, like: "Once again results were in the expected direction with reduced difficulties for children who were breastfed"</p> <p>Discussion</p> <p>First paragraph of the discussion, in my opinion should be put later in the text. You should start with an interpretation of your own results, than literature comparison, than talk about strengthens and limitation (here you could talk about the propensity score methodology), and finish with some final paragraphs talking about implication in a brief summary or conclusion</p> <p>Better interpretation of the results is needed, so ok in those who breastfed the CBCL showed a better score, but the number, the beta, is it that relevant what means 1,2 or 3 points. Is it really that relevant? If so, elaborate more.</p> <p>You said: "Given the types of behaviours where reductions for those breastfed are found in this study, and in the context of previous studies, a plausible hypothesis might be that of the nutrients found in breastmilk contributing to the growing infant's brain development. More research in this area, using well designed and rigorously sound methodology is first needed before firm conclusions can be drawn." I think your research did not evaluated this, and without a reference this statement seems more like an opinion of the authors, maybe rephrase or discuss it better with some literature.</p> <p>You need to deeply discuss this findings with the literature, your discussion is way too short, considering all the evidence there is regarding breastfeeding benefits, in terms of mental health outcomes, behavioural problems, and other related outcomes</p> <p>You do not discuss you losses, from almost 16 thousands you go to a little more than 4 thousands, selection bias is a possibility here, and you should discuss it.</p> <p>In general the discussion section should also be re-evaluated and maybe re-written, in order to elaborate more your ideas and really interpreted and compare your finding and maybe even rise new hypothesis.</p> |
|--|--|

|                        |   |
|------------------------|---|
| <b>REVIEWER</b>        | Stephanie D'Souza   |
|                        | COMPASS Research Centre University of Auckland, New Zealand |
| <b>REVIEW RETURNED</b> | 21-Aug-2018   |

|                         |  |
|-------------------------|--|
| <b>GENERAL COMMENTS</b> | <p>I commend the authors for the method used in their article. It is a more effective way of investigating potentially causal pathways of breastfeeding on child behavioural problems. However, more information on background literature and balance diagnostics is needed.</p> <p>Introduction</p> <p>In your introduction, as you talk about breastfeeding in relation to developed/developing countries, I would recommend giving a brief background on the socioeconomic status of Chile.</p> <p>You only mention the study by Girard et al. in passing. Given that replicating this study's findings is one of your objectives, more information on the study itself is needed.</p> <p>Minor revision: pg 4, line 4-5 - capitalise Lancet.</p> |
|-------------------------|--|

|  |  |
|--|--|
|  | <p><b>Methods</b><br/>On page 5, line 4-5, you state that "mothers who were still breastfeeding when behavioural measures were collected and who had breastfed more than six, but less than 12-months, were excluded". By behavioural measures, do you mean breastfeeding behaviour? This needs to be made clearer.</p> <p><b>Minor revisions:</b><br/>Pg 5, line 30-31 - change "on 3-point likert scale to "on a 3-point Likert scale".<br/>Pg 6, line 23-24 - I'm assuming that by age of child, you mean age at CBCL assessment? Just make this clearer.<br/>Pg 7, line 2 - I'm assuming that you're using the standardised statistical significance threshold of <math>p &lt; .05</math>, but I would recommend stating this explicitly.</p> <p><b>Results</b><br/>More information is needed on whether matching has resulted in similar distributions of covariates between those who were breastfed and who were not breastfed. For that reason, I would recommend including the standardised differences for each covariate.</p> <p>I would also recommend including a visualisation of the distribution of propensity scores for those who were and were not breastfed, and indicate the common area of support.</p> <p><b>Discussion</b><br/>The last sentence is awkwardly phrased - I would recommend rephrasing.</p> |
|--|--|

### VERSION 1 – AUTHOR RESPONSE

| EDITOR/REVIEWER COMMENTS  | AUTHOR'S RESPONSE   | REFERENCE PAGE             |
|---|---|----------------------------|
| <b>Editor Comments:</b>   |   |                            |
| 1. Please complete and include a STROBE checklist, ensuring that all points are included and state the page numbers where each item can be found. | We have now completed the STROBE checklist stating the page numbers where each item may be found.         | <b>Supplementary files</b> |
| 2. Please clearly mark the Introduction with a heading.   | We have now included a heading to clearly indicate the beginning of the introduction section.             | <b>Pg. 3</b>               |
| 3. Please re-upload your supplementary files in PDF format.   | Supplementary files have now been converted and uploaded in PDF format.                                   | <b>Supplementary files</b> |
| 4. Authors must include a statement in the methods section of   | We have now included the following statement on pg. X in line with the new journal requirements: "Patient | <b>Pg. 6</b>               |

|   |   |                          |
|---|---|--------------------------|
| <p>the manuscript under the sub-heading 'Patient and Public Involvement'.</p>   | <p>and Public Involvement: The development of the research question and outcome measures, along with study design and recruitment to, were not directly informed by patients' priorities, experience or preference. Study findings will be disseminated to the Ministry of Labor and Social Welfare, whom were responsible for waves 1 and 2, and the Ministry of Social Development, who is currently responsible for wave 3 of the ELPI cohort, ensuring greater likelihood of dissemination to study participants."</p>  |                          |
| <p><b>Reviewer 1 Comments:</b></p>  |   |                          |
| <p>1. First, I would like to congratulate the authors for the selected research subject. It is of interest, and an important addition to the literature and what it is known of the long term effects of breastfeeding.</p>   | <p>We thank the reviewer for their comment.</p>   | <p><b>No change.</b></p> |
| <p>2. The introduction could be a little more straight forward as well as the objective presented, you could explain your methodology in the methods section, I think you do not need to make a preview of it in the objective, and maybe discuss those points in the discussion section, and discuss the advantages, disadvantages and limitations of this method.</p> | <p>We have carefully considered this suggestion by the reviewer. We are inclined to keep the structure of the introduction and objectives as are, as we believe that it is important to highlight the methods used within the objectives, despite, as the reviewer correctly states, this being part of the methodology. For example, we believe that using this approach helps to diminish selection bias between groups on observables, getting us closer to potential 'causal' inferences and thus, is appropriately situated within the objectives section. We also believe that it is important to discuss the matching choices made within the methods section so that there is a clear understanding of the pros/cons of our choices, supported by previous literature, when reading through and evaluating the results.</p> | <p><b>No change.</b></p> |
| <p>3. First paragraph of the methods section is clear until you start talking about the 4300 children who were included in the study, you said it</p>   | <p>Yes, this is correct, our inclusion criteria were with respect to having complete data on all confounding variables to be used in matching at wave 1. In line with the reviewer's comment, we have now tried to make this clearer: "Inclusion criteria</p>   | <p><b>Pg. 5</b></p>      |

|   |   |                     |
|---|---|---------------------|
| <p>were all who had data on the confounding, is this still in wave one? It was clear after 3-4 times of reading it that actually it was the same wave, and from the 15000 only this 4thousand had all the need data. Maybe rephrase a little to make it clearer.</p>          | <p>in this study were children aged seven to 24-months, who had complete data on all confounders at wave one, and who were born full term (n=4,375). Additionally, mothers who were still breastfeeding when behavioural measures were collected and who had breastfed more than six, but less than 12-months, were excluded (n=442), as it was not possible to identify whether they should be included in the group of children breastfed between seven and 12-months or in the extended breastfeeding group. This resulted in a possible sample of 3,933 children and their families, 50.6% of whom were boys (n=1,992) at wave one. However, missing outcome data (i.e., child behaviours) at wave two in 2012, resulted in a final sample 3,037”</p> |                     |
| <p>4. It is unclear until this point if this a longitudinal or cross sectional study, maybe you should start the paragraph stating this.</p>  | <p>In line with the reviewer’s suggestion, we have now specified that this is a longitudinal study in the objectives: “To examine breastfeeding and children’s behavioural outcomes longitudinally, using a quasi-experimental statistical technique to reduce observable differences between groups, whereby attempting to address inherent limitations in observational studies”. Additionally, in the methods section we state the following: “Families recruited in the second wave (i.e., in 2012) were not considered in this study given that child outcomes were not available longitudinally”.</p>   | <p><b>Pg. 4</b></p> |
| <p>5. Along the methods section, you start giving some explanations om why you used this method, or what literature says about the confounder, or ethical issues, this should be discuss in the discussion section, in the methods you should only describe what you did.</p> | <p>We have removed: “Given ethical issues with randomisation” in the beginning paragraph of the statistical analysis section in line with the reviewer’s suggestion. However, given the plethora of methods for which to conduct PSM, we feel that the methods section is the appropriate place, at first introduction, to understand why we made the choices we did and how they are supported by the literature before going on to reading the results.</p>   | <p><b>Pg. 6</b></p> |
| <p>6. You do not mention the means correlation you made</p>   | <p>We included a statement detailing that the bivariate correlations, along with the means and standard</p>   | <p><b>Pg. 5</b></p> |

|   |  |  |
|---|--|--|
| <p>with the CBCL scores. In addition, in my experience the CBCL normally is not normally distributed, therefore a mean comparison of correlation is not adequate.</p>                                     | <p>deviations of child behaviours are presented in Table 2 within the methods section: “Means and standard deviations, along with correlations between subscales are presented in Table 2”. The correlations are merely presented as a descriptive analysis of the associations between behaviours within the sample, the correlations are not used as a comparison between groups.</p>  |  |
| <p>7. Same as the for the regression model used, comparing means is not adequate for the CBCL unless you demonstrated that the models complied to the basic requirements of a regression model.</p>       | <p>Regression analysis were only used with covariates on breastfeeding to create the propensity score for each child, with the selection of included covariates theoretical driven. We did not use regression analysis for the CBCL. For the t-tests, given the large cohort used, the normality assumption is assumed under the central limit theorem. Simulation studies have also found validity of t-tests in non-normally distributed data (e.g., Lumley T, Diehr P, Emerson S, Chen L. The importance of the normality assumption in large public health data sets. Annual review of public health. 2002 May;23(1):151-69).</p>  |  |
| <p>8. Results section should be re-written, it is not informative. Talking about being or not significant is not as relevant as telling us the amount of difference and correctly showing this Betas.</p> | <p>We fully appreciate the issue raised here by the reviewer. The limited journal space however, creates challenges. We have thus made the decision to present the results of the analysis mostly within Table format and not to repeat all of it within the text of the results section in order to save space for writing elsewhere. For all statistically significant results, we do outline the difference scores within the results section and we have now also included the effect size for each statistically significant finding to provide more context: “These results remained significant following matching whereby children who were breastfed had lower scores on these subscales (i.e., a mean difference of -1.00, <math>d = -0.23</math> and -1.02, <math>d = -0.27</math> respectively)” and “After matching, significant differences remained for emotional reactivity and attention problems only (i.e., a mean difference of -0.86, <math>d = -0.21</math> and -0.50, <math>d = -0.22</math>, respectively), with reduced difficulties for children who</p> | <p><b>Pgs. 6-7 and Figures 1 &amp; 2</b></p> |

|  |  |                          |
|--|--|--------------------------|
|  | <p>were breastfed". In addition, and in line with suggestions of the second reviewer, we have now also included two additional Figures, which display the overlapping support in the distribution of the propensity scores, along with the standardized differences on all covariates, pre and post matching. Please see Figures 1 and 2: "All children fell within the area of common support which refers to cases being excluded as a result of not fitting within the specified caliper. See Figure 1 for the overlapping support of the distribution of propensity scores. To ensure the overall quality of the matching procedure, balance checks were conducted on individual confounders and the overall models. For individual factors, remaining bias ranged between 0.0 and 18.8% (see Figure 2) and the overall mean remaining bias for models ranged between 5.5% and 7.2%. It has been suggested that less than 20% remaining bias is indicative of good matching,<sup>35</sup> thus we concluded that our matching was successful."</p> |                          |
| <p>9. It is also lacking fundamental parts, like a description of our population in terms of the variables you used.</p>   | <p>Similarly to above, and due to the limited journal space, we have made the decision to present some results within Table format only, instead of within the text. Please see Table 1 for the description of the cohort used.</p>  | <p><b>Pg. 13</b></p>     |
| <p>10. Also interpretations and opinions should not be included, like: "Once again results were in the expected direction with reduced difficulties for children who were breastfed"</p>   | <p>The statements demonstrate that the results support the directional hypothesis made and stated within the objectives section. However, in line with the reviewer's suggestion, both statements have now been removed from the results section.</p>  | <p><b>Pg. 7</b></p>      |
| <p>11. First paragraph of the discussion, in my opinion should be put later in the text. You should start with an interpretation of your own results, than literature comparison, than talk about strengthens and limitation (here you</p> | <p>We have carefully considered the suggestion of the reviewer. This is however a stylistic preference and we are more inclined to keep the structure of the discussion as is.</p>   | <p><b>No change.</b></p> |

|  |  |                        |
|--|--|------------------------|
| <p>could talk about the propensity score methodology), and finish with some final paragraphs talking about implication in a brief summary or conclusion.</p>   |  |                        |
| <p>12. Better interpretation of the results is needed, so ok in those who breastfed the CBCL showed a better score, but the number, the beta, is it that relevant what means 1,2 or 3 points. Is it really that relevant? If so, elaborate more.</p> | <p>We thank the reviewer for raising this important point regarding the discussion of practical effects and not just statistical significance. In line with this, we have now calculated Cohen's <i>d</i> to provide effect sizes for each finding (inserted within the results): "These results remained significant following matching whereby children who were breastfed had lower scores on these subscales (i.e., a mean difference of -1.00, <i>d</i> = -0.23 and -1.02, <i>d</i> = -0.27 respectively)" and "After matching, significant differences remained for emotional reactivity and attention problems only (i.e., a mean difference of -0.86, <i>d</i> = -0.21 and -0.50, <i>d</i> = -0.22, respectively), with reduced difficulties for children who were breastfed.". We have now also added the following to the discussion section: "While our results suggest statistically significant differences in favour of children who were breastfed at least six full months (and up until 12 full months), as compared to those who were never breastfed on emotional reactivity, somatic complaints, and inattention, the magnitude of effect for each behaviour was found to be small (i.e., Cohen's <i>d</i> &lt; .30). The practical and clinical significance of our results is arguably interpretable in the eye of the 'stakeholder'. A small reduction in a child's emotional reactivity, somatic complaints, and/or inattention in everyday situations may carry greater importance to a multiparous or first-time mother experiencing high levels of stress and fatigue as a result of limited financial and/or personal resources. On the other hand, within a clinical context, the effect sizes found may be perceived as carrying less practical importance."</p> | <p><b>Pgs. 7-8</b></p> |

|   |  |                       |
|---|--|-----------------------|
| <p>13. You said: “Given the types of behaviours where reductions for those breastfed are found in this study, and in the context of previous studies, a plausible hypothesis might be that of the nutrients found in breastmilk contributing to the growing infant’s brain development. More research in this area, using well designed and rigorously sound methodology is first needed before firm conclusions can be drawn.” I think your research did not evaluated this, and without a reference this statement seems more like an opinion of the authors, maybe rephrase or discuss it better with some literature.</p> | <p>The reviewer is indeed correct, we did not directly evaluate this in our study. We have now included references to support our hypothesis of potential mechanisms underlying our results: ‘Given the types of behaviours where reductions for those breastfed are found in this study, and in the context of previous studies (e.g.,8-10, 36-37), a plausible hypothesis might be that of the nutrients found in breastmilk contributing to the growing infant’s brain development’, prior to stating that ‘More research in this area, using well designed and rigorously sound methodology is first needed before firm conclusions can be drawn.’</p>   | <p><b>Pg. 8</b></p>   |
| <p>14. You need to deeply discuss this findings with the literature, your discussion is way too short, considering all the evidence there is regarding breastfeeding benefits, in terms of mental health outcomes, behavioural problems, and other related outcomes.</p>  | <p>We agree with the reviewer that the state of findings within the literature extends beyond the scope of our discussion. Once again, this has been a challenging issue due to space limitations and providing a balance between a discussion regarding our specific results, anchored of course to the specific body of findings for breastfeeding and behaviour, rather than a review of all of the many benefits of breastfeeding on development. We are open to extending our discussion further if the journal editor agrees to allowing more space for this. That said, we have included additional discussion to address specific points raised earlier by yourself with respect to practical implications and limitations, in addition to in response to reviewer 2 (please see our direct responses below, in particular point 9).</p> | <p><b>Pg. 8-9</b></p> |
| <p>15. You do not discuss you losses, from almost 16 thousands you go to a little</p>   | <p>We thank the reviewer for raising this important point. Within the article summary in the limitations we do state the following: “As a</p>  | <p><b>Pg. 9</b></p>   |

|   |  |                     |
|---|--|---------------------|
| <p>more than 4 thousands, selection bias is a possibility here, and you should discuss it.</p>  | <p>result of the inclusion/exclusion criteria, the sample size was reduced from the entire cohort” but we did not initially add the implication of this due to the word limit. We have now also included the following within the limitations section “Due to our inclusion/exclusion criteria, the sample size was significantly reduced, with some statistically significant differences between the originally recruited cohort and those included in the current study, indicative of potential selection bias. Thus, warranting replication”, in addition to presenting the differences between the sample used and the entire cohort in the online supplement 1.</p> |                     |
| <p><b>Reviewer 2 Comments:</b></p>  |  |                     |
| <p>1. I commend the authors for the method used in their article. It is a more effective way of investigating potentially causal pathways of breastfeeding on child behavioural problems. However, more information on background literature and balance diagnostics is needed.</p> | <p>We thank the reviewer for their comment. We have now included more background literature and more information on balance diagnostics as requested. Please see the specific additions detailed below in response to the itemized comments.</p>   |                     |
| <p>2. In your introduction, as you talk about breastfeeding in relation to developed/developing countries, I would recommend giving a brief background on the socioeconomic status of Chile.</p>  | <p>In line with the reviewer’s suggestion, and in the context of space limitations, we have now included the following: “While economic growth has been observed, social inequalities in Chile remain high, particularly for women<sup>25</sup>.”</p>  | <p><b>Pg. 4</b></p> |
| <p>3. You only mention the study by Girard et al. in passing. Given that replicating this study’s findings is one of your objectives, more information on the</p>   | <p>In line with the reviewer’s suggestion, and in the context of space limitations, we have now included the following: “Moreover, we examined whether in using a Chilean cohort, we could replicate the findings of Girard et al.<sup>11-12</sup> regarding reduced hyperactivity for children breastfed, following</p>   | <p><b>Pg. 4</b></p> |

|  |  |                              |
|--|--|------------------------------|
| study itself is needed.  | propensity score matching, in two separate longitudinal Irish cohorts.”  |                              |
| 4. Minor revision: pg 4, line 4-5 - capitalise Lancet.   | We thank the reviewer for catching this. Lancet has now been capitalized.  | <b>Pg. 4</b>                 |
| 5. On page 5, line 4-5, you state that "mothers who were still breastfeeding when behavioural measures were collected and who had breastfed more than six, but less than 12-months, were excluded". By behavioural measures, do you mean breastfeeding behaviour? This needs to be made clearer. | We thank the reviewer for catching this. Behavioural measures is a typo, this should have stated at wave 1 when all covariates were collected. We have now specified this to ensure more clarity:<br>“Additionally, mothers who were still breastfeeding at wave one who had breastfed more than six, but less than 12-months, were excluded (n=442), as it was not possible to identify whether they should be included in the group of children breastfed between seven and 12-months or in the extended breastfeeding group.” | <b>Pg. 5</b>                 |
| 6. Pg 5, line 30-31 - change "on 3-point likert scale to "on a 3-point Likert scale".  | In line with this suggestion, we have now capitalized Likert:<br>“Parents rate each individual behavioural item on 3-point Likert scale...”  | <b>Pg. 5</b>                 |
| 7. Pg 6, line 23-24 - I'm assuming that by age of child, you mean age at CBCL assessment? Just make this clearer.  | Child age refers to age at wave 1 as children ranged from 7-24 months. We have now added:<br>“...and age at first assessment in wave 1”  | <b>Pg. 6</b>                 |
| 8. I'm assuming that you're using the standardised statistical significance threshold of $p < .05$ , but I would recommend stating this explicitly.  | In line with the reviewer's suggestion, we have added the following: “We use the term significant henceforth to denote statistical significance, using a threshold of $p = <.05$ .”.   | <b>Pg. 7</b>                 |
| 9. More information is needed on whether matching has resulted in similar distributions of covariates between those who were breastfed and who were not breastfed. For that reason, I would recommend including the  | In line with the reviewer's suggestion, we have now included in Figure 2, the standardized differences in remaining bias for each individual covariate pre and post matching: “For individual factors, remaining bias ranged between 0.0 and 18.8% (see Figure 2) and the overall mean remaining bias for models ranged between 5.5% and 7.2%.” Visual representation revealed more  | <b>Pgs. 8-9 and Figure 2</b> |

|   |   |                                    |
|---|---|------------------------------------|
| <p>standardised differences for each covariate.</p> | <p>clearly that while matching was particularly successful for the first two models (i.e., up to 6 months and between 7 and 12 months), the matching was less successful for the extended group as this group had very similar family and maternal level characteristics to those who were not breastfed. We have now also raised this within the discussion and limitations with respect to interpretation of our findings for the extended breastfeeding group: “Of interest and as can be seen in Table 1, mothers who breastfed for extended durations in Chile had similar characteristics to mothers who had never breastfed, lending to poorer quality matching. For example, in both the never- and extended breastfeeding groups, a significantly higher proportion of mothers had never worked, were in the public tier of the health system, had only completed education at the primary level and had below average scores on both the digit and vocabulary scales of the WAIS; factors which when previously controlled, have reduced observed associations between breastfeeding and children’s cognitive and behavioural development outcomes.” And “Relatedly, the quality of matching for the extended breastfeeding families as compared to the never breastfeeding families was not as successful compared to the matching between the other groups, due to the initial similarities on health and social factors. The included covariates used for matching were theoretically motivated and thus, we kept the integrity of matching variables intact across all models. However, the findings from this model (i.e., the extended breastfeeding families) warrants caution in interpretation. Future studies are needed to more carefully evaluate extended breastfeeding and potential associations with behavioural outcomes, in the context of differing confounding structure.”. We thank the reviewer for their suggestion.</p> |                                    |
| <p>10. I would also recommend including a</p>       | <p>In the statistical analysis section, we specified: “All children fell within the area of common support which</p>  | <p><b>Pg. 6-7 and Figure 1</b></p> |

|  |  |              |
|--|--|--------------|
| visualisation of the distribution of propensity scores for those who were and were not breastfed, and indicate the common area of support. | refers to cases being excluded as a result of not fitting within the specified caliper". In line with the reviewer's suggestion regarding a visualisation of the distribution, we have now also included this information in Figure 1: "See Figure 1 for the overlapping support of the distribution of propensity scores."  |              |
| 11. The last sentence is awkwardly phrased - I would recommend rephrasing.   | In line with this suggestion, we have now rephrased into two sentences: "A comprehensive answer to the question of effects on psychosocial development remains unanswered without the use of RCTs. However, with replication across regions, whilst using more stringent methodological approaches to help in reducing bias inherent in observational studies, promise for better understanding of potential mechanisms is viable" | <b>Pg. 9</b> |

#### VERSION 2 – REVIEW

|                        |   |
|------------------------|---|
| <b>REVIEWER</b>        | Christian Loret de Mola<br>Federal University of Rio Grande |
| <b>REVIEW RETURNED</b> | 13-Oct-2018   |

|                         |  |
|-------------------------|--|
| <b>GENERAL COMMENTS</b> | <p>First, I would like to congratulate the authors for this manuscript, the theme is relevant and clearly not well explored in the literature. I would only have to make few comments and suggestions</p> <p>1. Why did the authors use the health system as a proxy of income? This must be better explained in the manuscript. How is it that this variable could represent something like income in the Chilean context? I understand the context since and know it in part, however most readers would not. In addition, even when it could be a proxy of socioeconomic position, talking about income is kind of tricky in terms of economics, since it represents capacity of acquiring things at the moment. And being in a health plan might not necessarily represent that.</p> <p>2. Using means for your outcome seems a little inadequate to me, since the scores are most likely not normal, therefore comparing means is not adequate. Please make sure that all the assumptions of linear regression are met, and please show them as supplementary material.</p> |
|-------------------------|--|

|                        |   |
|------------------------|---|
| <b>REVIEWER</b>        | Stephanie D'Souza<br>COMPASS Research Centre, University of Auckland, New Zealand |
| <b>REVIEW RETURNED</b> | 10-Oct-2018   |

|                         |   |
|-------------------------|---|
| <b>GENERAL COMMENTS</b> | In my initial review of this manuscript, I felt that the paper was lacking in information on background literature and balance diagnostics. I now feel that the authors have adequately addressed the background literature, given the word limit restrictions. |
|-------------------------|---|

|  |   |
|--|---|
|  | <p>Additionally, I am pleased to see that the authors have included the standardised differences and acknowledged the limitations with matching for the extended breastfeeding group. The authors have also satisfactorily acknowledged the study's limitations. Overall, the authors should be commended on their use of advanced methodology to address their research question and I believe that the article is now acceptable for publication.</p> |
|--|---|

### VERSION 2 – AUTHOR RESPONSE

| EDITOR/REVIEWER COMMENTS  | AUTHOR'S RESPONSE  | REFERENCE PAGE    |
|---|--|-------------------|
| <b>Editor Comments:</b>   |  |                   |
| 5. Could the authors provide a range of values (minimum and maximum) for each subscale?   | The minimum and maximum values for each subscale have now been included in Table 2.  | <b>Pg. 13</b>     |
| 6. There is a recent systematic review the authors should cite:<br>Rev Saude Publica. 2018 Feb 5;52:9. doi: 10.11606/S1518-8787.2018052000439.<br>Breastfeeding and behavior disorders among children and adolescents: a systematic review. <a href="https://www.ncbi.nlm.nih.gov/pubmed/29412376">https://www.ncbi.nlm.nih.gov/pubmed/29412376</a> | This suggested reference has now been included on pg. 3.   | <b>Pg. 3</b>      |
| <b>Reviewer 1 Comments:</b>   |  |                   |
| 16. First, I would like to congratulate the authors for this manuscript, the theme is relevant and clearly not well explored in the literature.   | We thank the reviewer for their comment.   | <b>No change.</b> |
| 17. Why did the authors use the health system as a proxy of income? This must be better explained in the manuscript. How is it that this variable could represent something   | The reviewer raises an important point and one in which more context is admittedly needed for the reader. We have now included the following on pg. 6: "To note, the quality of services offered in the private and public healthcare system in Chile differ vastly, with higher quality services offered in | <b>Pg. 6</b>      |

|   |  |   |
|---|--|---|
| <p>like income in the Chilean context? I understand the context since and know it in part, however most readers would not. In addition, even when it could be a proxy of socioeconomic position, talking about income is kind of tricky in terms of economics, since it represents capacity of acquiring things at the moment. And being in a health plan might not necessarily represent that.</p> | <p>the private system; subsequently translating into a high cost of belonging to the private system. Moreover, for those employed, the tier of the healthcare system in which one belongs is directly related to income earnings, whereby employers pay into the healthcare system on their employees behalf, which is a calculated monthly percentage deductible based on individual income". The point about acquisition of things is well taken. However, it could equally be argued that the acquisition of access to better levels of service/treatment in healthcare would be a representation of the capacity to acquire 'things'.</p>  |   |
| <p>18. Using means for your outcome seems a little inadequate to me, since the scores are most likely not normal, therefore comparing means is not adequate. Please make sure that all the assumptions of linear regression are met, and please show them as supplementary material.</p>  | <p>We take on board the reviewer's concerns. Previous simulation studies have demonstrated validity of t-tests with non-normally distributed data (e.g., Lumley T, Diehr P, Emerson S, Chen L. The importance of the normality assumption in large public health data sets. Annual review of public health. 2002 May;23(1):151-69), in large cohorts such as in the current study. However, in light of the potential issues that may arise with hypothesis testing with non-normally distributed data, we have standardized all behaviour subscales and as a sensitivity analysis, we then re-ran all of our models (table is attached below). As can be seen, this had no impact on our hypothesis testing and all our results remained the same. As a result of the added complexity and less intuitive interpretation for the reader, in using differences in units of standard deviations rather than mean differences, we have kept our original mean differences analysis in the main paper. If the editor is keen on also including the sensitivity analysis using the standardized subscales as an online supplement, we have no objection.</p> | <p><b>Please see directly below. Can be included as an online supplement as well.</b></p> |

**Sensitivity Analysis: Breastfeeding and Children's Behavioural Problems Using Standardised Scores: Pre and Post Matching Results**

| Up to 6 months                 | Pre Matching |      |                 |      | Post Matching |      |                |      |
|--------------------------------|--------------|------|-----------------|------|---------------|------|----------------|------|
|                                | T            | C    | Diff (Sig.)     | S.E  | T             | C    | Diff (Sig.)    | S.E  |
| Emotionally reactive           | 0.01         | 0.29 | <b>-0.27**</b>  | 0.10 | 0.01          | 0.36 | <b>-0.34*</b>  | 0.14 |
| Anxious/depressed              | 0.01         | 0.11 | -0.09           | 0.09 | 0.01          | 0.08 | -0.07          | 0.12 |
| Somatic complaints             | -0.00        | 0.24 | <b>-0.24*</b>   | 0.09 | -0.00         | 0.38 | <b>-0.38**</b> | 0.14 |
| Withdrawn                      | 0.03         | 0.10 | -0.06           | 0.10 | 0.03          | 0.09 | -0.06          | 0.13 |
| Sleep problems                 | 0.01         | 0.09 | -0.07           | 0.09 | 0.01          | 0.21 | -0.20          | 0.13 |
| Attention problems             | 0.03         | 0.16 | -0.13           | 0.09 | 0.03          | 0.06 | -0.03          | 0.11 |
| Aggression                     | 0.02         | 0.14 | -0.11           | 0.10 | 0.02          | 0.11 | -0.08          | 0.14 |
| <b>Between 7 and 12 months</b> |              |      |                 |      |               |      |                |      |
| Emotionally reactive           | -0.08        | 0.29 | <b>-0.38***</b> | 0.09 | -0.08         | 0.20 | <b>-0.29*</b>  | 0.14 |
| Anxious/depressed              | -0.06        | 0.11 | -0.18           | 0.10 | -0.06         | 0.02 | -0.08          | 0.12 |
| Somatic complaints             | -0.05        | 0.24 | <b>-0.29**</b>  | 0.10 | -0.05         | 0.21 | -0.26          | 0.14 |
| Withdrawn                      | -0.09        | 0.10 | <b>-0.19*</b>   | 0.09 | -0.09         | 0.06 | -0.15          | 0.12 |
| Sleep problems                 | -0.05        | 0.09 | -0.14           | 0.09 | -0.05         | 0.04 | -0.09          | 0.13 |
| Attention problems             | -0.07        | 0.16 | <b>-0.24*</b>   | 0.09 | -0.07         | 0.17 | <b>-0.25*</b>  | 0.11 |
| Aggression                     | -0.07        | 0.14 | <b>-0.22*</b>   | 0.09 | -0.07         | 0.02 | -0.09          | 0.13 |
| <b>13 months or more</b>       |              |      |                 |      |               |      |                |      |
| Emotionally reactive           | 0.03         | 0.29 | <b>-0.26*</b>   | 0.10 | 0.03          | 0.18 | -0.15          | 0.16 |
| Anxious/depressed              | 0.03         | 0.11 | -0.08           | 0.10 | 0.03          | 0.06 | -0.03          | 0.14 |
| Somatic complaints             | 0.03         | 0.24 | <b>-0.20*</b>   | 0.10 | 0.03          | 0.12 | -0.09          | 0.15 |

|                    |      |      |       |      |      |       |       |      |
|--------------------|------|------|-------|------|------|-------|-------|------|
| Withdrawn          | 0.03 | 0.10 | -0.06 | 0.10 | 0.03 | 0.10  | -0.07 | 0.14 |
| Sleep problems     | 0.03 | 0.09 | -0.05 | 0.10 | 0.03 | -0.06 | 0.10  | 0.15 |
| Attention problems | 0.02 | 0.16 | -0.14 | 0.10 | 0.02 | 0.02  | -0.01 | 0.12 |
| Aggression         | 0.02 | 0.14 | -0.11 | 0.10 | 0.02 | -0.14 | 0.17  | 0.15 |

Note: \*\*\* denotes significance at the  $p < .001$  level, \*\* at the .01 level, \* at the .05 level. T denotes 'treatment' (breastfed) and C denotes 'control' (not breastfed). 'Diff' represents the difference in scores between groups in units of standard deviations. S.E. refers to the standard errors. For being breastfed up to 6 months: N for the treatment group was 949 and 110 for the control group. For being breastfed between 7 and 12 months: N for the treatment group was 946 and 110 for the control group. For being breastfed 13 months or more: N for the treatment group was 1,006 and 110 for the control group.