## PEER REVIEW HISTORY

BMJ Paediatrics Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

This paper was submitted to a another journal from BMJ but declined for publication following peer review. The authors addressed the reviewers' comments and submitted the revised paper to BMJ Paediatrics Open. The paper was subsequently accepted for publication at BMJ Paediatrics Open.

## ARTICLE DETAILS

| | |
|---|---|
| **TITLE (PROVISIONAL)** | Probiotics, prematurity and neurodevelopment: Follow-up of a randomised trial |
| **AUTHORS** | Jacobs, Susan; Hickey, Leah; Donath, Susan; Opie, Gillian; Anderson, Peter; Garland, Suzanne; Cheong, Jeanie |

## VERSION 1 - REVIEW

| | |
|---|---|
| **REVIEWER** | Fleming, Paul<br>Homerton University Hospital<br>Homerton Row LONDON E9 6SR<br>Competing interests: I worked at the Royal Women's Hospital in 2008 as a neonatal fellow. The lead author and second author of this study are known to me. I have previously co-authored a paper on a different subject matter with the lead author. Between 2010-2013 I conducted a series of nested studies within the United Kingdom Probiotics in Pretem Babies (PiPS) study which was a competing study to the one referred to in this paper. |
| **REVIEW RETURNED** | 04-Jul-2017 |

| | |
|---|---|
| **GENERAL COMMENTS** | One of the most frequently stated concerns regarding probiotic use in preterm babies relates to their potential long term safety. This is the biggest and most comprehensive study of neurodevelopmental follow-up of babies recruited to one of the largest randomised controlled trials of probiotic use in preterm babies.<br><br>The importance of this paper cannot be overstated; the results add significantly to the field of probiotic research in preterm infants; the study design is appropriate; and the manuscript is exceedingly well written and presented.<br><br>I have no criticisms or comments other than to slightly contest the opening statement that probiotics confer a convincing impact on reducing late onset sepsis in preterm infants. Their effects on NEC and mortality are well established in the published literature.<br><br>Given these comments, my recommendation is to accept the paper without the need for revision. |

| | |
|---|---|
| **REVIEWER** | Vermeulen, Marijn<br>ErasmusMC, Rotterdam, The Netherlands<br>Competing interests: none |
| **REVIEW RETURNED** | 29-Jul-2017 |

| | |
|---|---|
| **GENERAL COMMENTS** | In the manuscript under review, the authors present the results of a follow-up study of an earlier (2007-2011) performed db-RCT on probiotics in preterm infants (<32 weeks, <1500 grams). A combination of B. infantis, S. thermophilus and B. lactis was administered until hospital discharge or term corrected age as compared to placebo in one of the largest RCTs in this patient population up till now. The primary outcome of this former RCT was culture-proven late onset sepsis, which was not reduced in the probiotic group as compared to the placebo group. One of the, over twenty, secondary outcomes of that RCT, was necrotizing enterocolitis, which was reduced in the treatment group, while mortality and NEC-free survival was not different. In their follow-up study the authors focus on the neurodevelopmental outcome at age 2-5 years, as measured by different validated tests. They conclude that survival without major disability in ex-preterms treated with probiotics is not different from those who received placebo.<br><br>The strength of this study is the clinical relevance of a subject globally discussed, as a great part of probiotic criticism is based on the lack of evidence for long term safety. We like to compliment the authors with the important work they have done on this subject, especially realizing that follow-up studies take a lot of effort and often show negative findings.<br><br>However, before acceptance for publication we have several suggestions for revision, that we think need to be addressed.<br><br>1. The most important improvement of the paper may be reached by reformulation of the conclusions. We strongly suggest more cautious interpretation and less generalizing of the results. This was properly done in the earlier Pediatrics paper, in which the authors discuss (and explicitly name) that they can only draw conclusions on the specific mixture (and dosage/setting/etc) of the probiotics tested. In the current manuscript, they draw several conclusions (in the abstract and in the last part of the discussion) that are far too general, by talking about “the safety of probiotics”. We think this does not help the delicate discussion on probiotics in preterms. Please realise that for the readers it may not be logical why probiotics should be administered if it does not reduce costs, nor improves outcome.<br>Both sentences in the conclusion of the abstract need more specification on the product used, to be valid. This study concerns only one product; not all probiotics! The same counts for the conclusions at the end of the paper.<br><br>2. We suggest that the introduction could benefit from a more logical structure. First, the authors clearly explain that NEC reduction or influencing the brain gut axis may improve neurodevelopment. Then, they state that the aim of the study is to prove safety (no harmful effects), of which the underlying logic or mechanisms are not discussed (why would it not be safe?), and indeed formulate a hypothesis that the probiotics have no harmful effects. This can also lead to confusion in the cost analysis, discussed later.<br><br>3. Follow-up was not planned at design of the study, but was already announced in their first paper. A sample size calculation was apparently not performed, we assume because they had to deal with a convenience sample anyway. However, some calculation to get an impression on the power, and on what followup rate was required to |

| | |
|---|---|
| | detect a clinical significant difference (or no difference; how would they define the cutoff?) may be helpful.<br><br>4. The authors describe that the follow-up rate is far from optimal, which we understand from a practical point of view. This may, however, impact the validity of the study. In our opinion, this is not discussed enough. We agree with the authors that multiple imputation of the large amount of missing information is debatable with little new information on the cohort. However, we think the authors have showed little attempt to judge potential bias due to incomplete data. The scientific quality of the manuscript would improve if the authors would perform sensitivity analysis, preferably a tipping point analysis, to evaluate the potential impact of missing not at random. This can help the readers to judge the clinical plausibility of the extremes that would change your conclusions (in both directions: in favor of probiotics or in favor of placebo).<br><br>5. The neurodevelopmental tests used, are applied in a wide age-range from 2 to 5 years. Therefore, different tests needed to be used. We think that this may have increased the variation and may have interfered with the results. Testing all children at the same age with the same test would have been preferable, which has not been discussed. Also, for the competing outcome of death this may be relevant; how high was mortality between 2 and 5 years in both groups? How do the authors think the age-range may have influenced their conclusions? What do they suggest for future follow up studies?<br><br>6. Page 5, line 40 states: The pre-specified primary outcome was “survival free of major neurosensory impairment. But the authors do not specify what the observation period is: do the authors mean: survival until testing date? What factors determined the age of testing? Is their potential bias in this design, or did the authors try to limit that somehow?<br><br>7. The limited sample size raises also questions on the number of tests performed (Table 4 shows twenty-five p-values, while using a cut off of <0.05 for statistical significance). The authors do mention the risk of false positive findings in the discussion on the hearing loss, but have not suggested how they could have corrected for multiple testing.<br><br>8. The hearing loss, indeed, may be a chance finding. But if not, how would the authors explain the finding? We appreciate the additional analysis on the antibiotics, but the question remains. Any speculations on the mechanisms possible? Or do the authors stick with the suggestion to firstly confirm the finding in other studies?<br><br>9. The cost-effectiveness part, that focused on the period until 24 months corrected age, is of great interest and relevance but seems to be forced into this manuscript. The methods of the cost-analysis are poorly described, and the results are not discussed. In the introduction it is not clear why the authors expect/hypothesize a reduction in health costs. The earlier RCT only showed a reduction in NEC; if that would lead to cost reduction during hospital stay (did it?) then no 2 year follow up was necessary. By focusing on 2 years, they seem to expect improved neurodevelopment in the probiotics group, which was not their primary hypothesis. As an average reader is not familiar with the Australian Medicare system it is unclear what kind of costs are included: primary care, teaching |

| | |
|---|---|
| | support, social costs etc? The authors now leave us with questions like: Why are the costs comparable if NEC is reduced in one of the groups? How do they explain this? In other words, without extra explanation and discussion this part creates more questions than it answers and I would prefer to leave it out of this paper and write a separate paper on this subject.<br><br>10. Abstract: without explaining that you found a reduction in NEC in the former study it is not logical to introduce that (NEC reduction) in the conclusions. We suggest to leave the last sentence out. |

| | |
|---|---|
| **REVIEWER** | Berrington, Janet<br>Newcastle University and Newcastle Hospitals, Newcastle, UK<br>Competing interests: None |
| **REVIEW RETURNED** | 31-Jul-2017 |

| | |
|---|---|
| **GENERAL COMMENTS** | As noted by the authors the major drawback is that this was not planned at the beginning of the study. This unfortunately impacts both on interpretation but also presentation of the data, as it is written very much from the primary study perspective throughout. This means that things that would otherwise be fundamental to writing this paper have been 'missed' or are presented unusually. For example, the methods section does not carry any information about ethics relating to this study (it is in the outcome section), about how parents were approached to consider taking part in the follow up study (mail/phone), by whom (local PI or study CI), how many times they were approached before being considered 'lost' to follow up, why the assessment was targeted at 2.5 years, who actually did the ND assessments (local team or study specific team), whether local data was sought if the study team did not make contact (I assume that local f/u was being done in these infants), what the power of the study was thought to be at the front end, etc. I am sure the authors have all this information and that it can be added, but as I read the paper it felt like an add on to pro-prems rather than an important well presented study in it's own right.<br>Other similar comments are below:<br>Abstract: Main outcome measure is actually 'survival free of major neurodevelopmental impairment', and the conclusion needs to be specific to this probiotic. Not sure that NEC reduction should feature in the abstract as not this study.<br>Health care costs - I am not sure how this would be expected to be measured in Australia, but in the UK a study attempting to measure costs would need to do much more than has been done here, where it appears that the hospital 'bill' has been added to those from a 'caregiver' report, but as things might be different in Australia, I do not claim to know enough about this to know whether this is legitimate or not. Personally I would be inclined to remove the parts related to costs from this paper.<br>Timing of assessments: these are presented as means and SD but median and range feels more informative - I can see that some infants were >4 years when they were assessed but can't work out quite how old the oldest was. Was there a cut-off above which assessments were no longer sought or not? Were all assessments done in one go? By whom? Why does Bayleys overestimate Australian infants??<br>Conclusions/hypotheses and some text are written unusually, for example P13 line 25 is oddly written, and the hypothesis is 'not |

| | |
|---|---|
| | negatively affect ND and reduce costs (which goes with improving ND rather than not making it worse), and the conclusion does not mention costs. throughout there is a tendency to refer to this data to support any or all probiotics, but the data pertain only to the specific probiotic used in ProPrems.<br>In the 'what is known' box (and in the conclusion) this should be 'some' probiotics reduce NEC.....<br>And in the 'what this study adds' the point on NEC should be removed as this data is ND outcome, not NEC |

**VERSION 1 – AUTHOR RESPONSE**

Reviewer: 1

*No revisions recommended.*

Reviewer: 2

**Comments to the Author**

1. The most important improvement of the paper may be reached by reformulation of the conclusions. We strongly suggest more cautious interpretation and less generalizing of the results. This was properly done in the earlier Pediatrics paper, in which the authors discuss (and explicitly name) that they can only draw conclusions on the specific mixture (and dosage/setting/etc) of the probiotics tested. In the current manuscript, they draw several conclusions (in the abstract and in the last part of the discussion) that are far too general, by talking about "the safety of probiotics". We think this does not help the delicate discussion on probiotics in preterms. Please realise that for the readers it may not be logical why probiotics should be administered if it does not reduce costs, nor improves outcome.
Both sentences in the conclusion of the abstract need more specification on the product used, to be valid. This study concerns only one product; not all probiotics! The same counts for the conclusions at the end of the paper.

*We have amended the abstract conclusions as follows:*
*"Adminstration of the probiotics combination Bifidobacterium infantis, Streptococcus thermophilus and Bifidobacterium lactis to very preterm babies from soon after birth until discharge home or term CA, did not affect survival free of major disability in early childhood. This probiotics mixture reduced necrotising enterocolitis without adverse effects on neurodevelopment or behaviour in very preterm children."*

*The conclusions at the end of the manuscript have also been amended as requested:*
*"In conclusion, administration of the probiotics combination Bifidobacterium infantis, Streptococcus thermophilus and Bifidobacterium lactis to very preterm infants from soon after birth until term corrected age did not negatively impact on survival free of neurodevelopmental impairment in early childhood. This probiotics mixture reduces necrotising enterocolitis in very preterm infants, with apparent safety with respect to early childhood development."*

2. We suggest that the introduction could benefit from a more logical structure. First, the authors clearly explain that NEC reduction or influencing the brain gut axis may improve neurodevelopment. Then, they state that the aim of the study is to prove safety (no harmful effects), of which the underlying logic or mechanisms are not discussed (why would it not be safe?), and indeed formulate a hypothesis that the probiotics have no harmful effects. This can also lead to confusion in the cost analysis, discussed later.

*We have removed the sentence about potentially harmful effects of probiotics on neurodevelopment, so that the stated aim of this follow-up study is clearer. It was not to "prove safety", but as stated in the manuscript was "to investigate the effects of the probiotics combination Bifidobacterium infantis, Streptococcus thermophilus and Bifidobacterium lactis on neurodevelopment and behaviour of a large cohort of very preterm participants in the ProPrems trial at 2-5 years of age, corrected for prematurity (CA)".*

3. Follow-up was not planned at design of the study, but was already announced in their first paper. A sample size calculation was apparently not performed, we assume because they had to deal with a convenience sample anyway. However, some calculation to get an impression on the power, and on what followup rate was required to detect a clinical significant difference (or no difference; how would they define the cutoff?) may be helpful.

*Although follow-up was not initially planned in the randomised controlled trial, this follow-up study did include the following sample size calculation as determined by the original study: "The ProPrems large sample size of 1,100 participants is adequately powered to demonstrate small but clinically significant differences between the groups. The table below illustrates the sample numbers expected assuming (i) mortality 6.5% (controls) 3.5% (intervention), based on the Deshpande et al 2010 meta-analysis which found 7.5% vs 3.5%. [3] Given the overall mortality rate in Proprems (5%) we have lowered the expected mortality rate in the controls (ii) Follow-up rate of 90% (this is a conservative estimate based on previous longitudinal studies undertaken by the CIs on this project) (iii) disability rates in survivors 20% (controls), 15% (intervention). [67]"*

4. The authors describe that the follow-up rate is far from optimal, which we understand from a practical point of view. This may, however, impact the validity of the study. In our opinion, this is not discussed enough. We agree with the authors that multiple imputation of the large amount of missing information is debatable with little new information on the cohort. However, we think the authors have showed little attempt to judge potential bias due to incomplete data. The scientific quality of the manuscript would improve if the authors would perform sensitivity analysis, preferably a tipping point analysis, to evaluate the potential impact of missing not at random. This can help the readers to judge the clinical plausibility of the extremes that would change your conclusions (in both directions: in favor of probiotics or in favor of placebo).

*Thank you for these important comments and the opportunity to include the results of the following sensitivity analysis in the manuscript: In order to assess the sensitivity of our results to possible differential survival free of major neurodevelopmental impairment rates in the study participants not included in our follow-up results, we investigated hypothetical rates of survival free of major neurodevelopmental impairment for these missing data in the two study arms in order to find the minimum difference which would have resulted in a p value < 0.05 from chi-squared test based on the entire cohort.*

*Rates of survival free of major neurodevelopmental impairment, chi-square p value < 0.05 for difference in disability-free survival between placebo and probiotics.*

| *Non-missing probiotic (N=373)* | *Non-missing Placebo (N=362)* | *Missing probiotic (N=175)* | *Missing placebo (N=189)* | *Total probiotic (N=548)* | *Total placebo (N=551)* |
|---|---|---|---|---|---|
| *75.3% (281)* | *74.9% (271)* | *<= 57.1% (<=100)* | *75.1% (142)* | *<=69.5% (381)* | *75.0% (413)* |
| *75.3% (281)* | *74.9% (271)* | *>=89.7% (>=157)* | *75.1% (142)* | *>=79.9% (438)* | *75.0% (413)* |
| *75.3% (281)* | *74.9% (271)* | *>=82.9% (>=145)* | *<=68.3% (<=129)* | *>=77.7% (>=426)* | *<=72.6% (<=400)* |
| *75.3% (281)* | *74.9% (271)* | *<=66.3% (<=116)* | *>=83.1% (>=157)* | *<=72.4% (<=397)* | *>=77.7% (>=428)* |

*Based on the simulations above, there would need to be a minimum difference of ~14% between rates of survival free of major neurodevelopmental impairment in missing data in order to obtain a p value < 0.05 for the total cohort.*

*The following has been added to the Results: "Given the high rates of loss to follow-up, we performed a sensitivity analysis to possible differential rates of disability free survival in the study participants who did not attend follow-up. In order to achieve a p value <0.05 for a difference of disability free survival rates between probiotics and placebo groups, we would have needed a difference of at least 14% between the groups who did not attend follow-up."*

*And the following has been added to the Discussion: "Following the sensitivity analyses, it is unlikely that a difference of disability free survival between probiotics and placebo who attended follow-up would be ≥14%. Therefore we do not believe our conclusions would be altered."*

5. The neurodevelopmental tests used, are applied in a wide age-range from 2 to 5 years. Therefore, different tests needed to be used. We think that this may have increased the variation and may have interfered with the results. Testing all children at the same age with the same test would have been preferable, which has not been discussed. Also, for the competing outcome of death this may be relevant; how high was mortality between 2 and 5 years in both groups? How do the authors think the age-range may have influenced their conclusions? What do they suggest for future follow up studies?

*A sensitivity analysis was performed excluding the 62 children who underwent assessment at over 42 months' corrected age and who were not assessed with the Bayley-III. This did not change the results of the study.*

*The following has been added to the discussion: "However, based on the sensitivity analysis, the age range is unlikely to have influenced the study conclusions because numbers were small (n=62) and also because dichotomous outcomes based on SD (2SD below the mean) were used to standardise the different tests. Future trials of probiotics should include neurodevelopmental follow-up outcomes from inception."*

6. Page 5, line 40 states: The pre-specified primary outcome was "survival free of major neurosensory impairment. But the authors do not specify what the observation period is: do the authors mean: survival until testing date? What factors determined the age of testing? Is their potential bias in this design, or did the authors try to limit that somehow?

*Ideally all survivors would have been assessed at 24 months' of age corrected for prematurity (CA). However, neurodevelopment was not a planned outcome of the ProPrems RCT and by the time funding and ethics approvals were obtained many ProPrems participants were already more than 2 years CA. Where possible, retrospective consent was obtained to access Bayley-III 2 year CA results. Where this was not possible, assessments were performed as soon as practicable. A small number (n = 62) were assessed above the Bayley-III ceiling when the WPPSI and Movement ABC were used.*

7. The limited sample size raises also questions on the number of tests performed (Table 4 shows twenty-five p-values, while using a cut off of <0.05 for statistical significance). The authors do mention the risk of false positive findings in the discussion on the hearing loss, but have not suggested how they could have corrected for multiple testing.

*We acknowledge that there are multiple comparisons and have interpreted results using patterns of changes and magnitude of differences (Relative Risks and 95% Confidence Intervals), not just absolute P-values.*

8. The hearing loss, indeed, may be a chance finding. But if not, how would the authors explain the finding? We appreciate the additional analysis on the antibiotics, but the question remains. Any speculations on the mechanisms possible? Or do the authors stick with the suggestion to firstly confirm the finding in other studies?

*The conclusion has been amended as follows: One possible explanation for this finding is that probiotics may reduce cochlear injury and therefore sensorineural hearing loss. Although only described to date in adult mice with Lactococcus lactis reducing age-related cochlear degeneration and hearing loss (Oike Sci Rep 2016), sensorineural hearing loss in preterm infants is also predominantly mediated by cochlear and outer hair-cell mediated injury (Cristobal ADC FN 2008). As this is only speculative,…….*

9. The cost-effectiveness part, that focused on the period until 24 months corrected age, is of great interest and relevance but seems to be forced into this manuscript. The methods of the cost-analysis are poorly described, and the results are not discussed. In the introduction it is not clear why the authors expect/hypothesize a reduction in health costs. The earlier RCT only showed a reduction in NEC; if that would lead to cost reduction during hospital stay (did it?) then no 2 year follow up was necessary. By focusing on 2 years, they seem to expect improved neurodevelopment in the probiotics group, which was not their primary hypothesis. As an average reader is not familiar with the Australian Medicare system it is unclear what kind of costs are included: primary care, teaching support, social costs etc? The authors now leave us with questions like: Why are the costs comparable if NEC is reduced in one of the groups? How do they explain this? In other words, without extra explanation and discussion this part creates more questions than it answers and I would prefer to leave it out of this paper and write a separate paper on this subject.

*We agree with the recommendation to remove costs from this manuscript and to report them in a separate paper.*

10. Abstract: without explaining that you found a reduction in NEC in the former study it is not logical to introduce that (NEC reduction) in the conclusions. We suggest to leave the last sentence out.

*This is a good point, thank you. We have changed the study design as follows to include NEC: Follow-up study of survivors of a double-blinded, placebo-controlled, randomised trial of probiotic effects on late-onset sepsis in very preterm infants that found reduced necrotising enterocolits (NEC).*

Reviewer: 3

**Comments to the Author**
Thank you for asking me to review this paper which seeks to add to the paucity of literature on the effect of probiotic receipt on longer term outcomes in preterm infants.
As noted by the authors the major drawback is that this was not planned at the beginning of the study. This unfortunately impacts both on interpretation but also presentation of the data, as it is written very much from the primary study perspective throughout.
This means that things that would otherwise be fundamental to writing this paper have been 'missed' or are presented unusually.
For example, the methods section does not carry any information about ethics relating to this study (it is in the outcome section), about how parents were approached to consider taking part in the follow up study (mail/phone), by whom (local PI or study CI), how many times they were approached before being considered 'lost' to follow up, why the assessment was targeted at 2.5 years, who actually did the ND assessments (local team or study specific team), whether local data was sought if the study team did not make contact (I assume that local f/u was being done in these infants), what the power of the study was thought to be at the front end, etc. I am sure the authors have all this information and that it can be added, but as I read the paper it felt like an add on to pro-prems rather than an important well presented study in it's own right.
Other similar comments are below:
Abstract: Main outcome measure is actually 'survival free of major neurodevelopmental impairment', and the conclusion needs to be specific to this probiotic. Not sure that NEC reduction should feature in the abstract as not this study.

*See reviewer 2, number 10.*

Health care costs - I am not sure how this would be expected to be measured in Australia, but in the UK a study attempting to measure costs would need to do much more than has been done here,

where it appears that the hospital 'bill' has been added to those from a 'caregiver' report, but as things might be different in Australia, I do not claim to know enough about this to know whether this is legitimate or not. Personally I would be inclined to remove the parts related to costs from this paper.

*Costs have been removed from this manuscript.*

Timing of assessments: these are presented as means and SD but median and range feels more informative - I can see that some infants were >4 years when they were assessed but can't work out quite how old the oldest was. Was there a cut-off above which assessments were no longer sought or not? Were all assessments done in one go? By whom? Why does Bayleys overestimate Australian infants??

*The following has been added for clarification: "Participants were assessed in a single session by a trained examiner. The Bayley-III over-estimates development, and as such, under-estimates developmental delay. This is not an Australian phenomenon, and is likely due to a mixed sampling procedure used for the Bayley-III standardization (Anderson PJ & Burnett A. (2016). Assessing developmental delay in early childhood – concerns with the Bayley-III scales. The Clinical Neuropsychologist, 31(2), 371-381)."*

Conclusions/hypotheses and some text are written unusually, for example P13 line 25 is oddly written, and the hypothesis is 'not negatively affect ND and reduce costs (which goes with improving ND rather than not making it worse), and the conclusion does not mention costs. throughout there is a tendency to refer to this data to support any or all probiotics, but the data pertain only to the specific probiotic used in ProPrems.

*Thanks you, these have been rectified in the revised manuscript.*

In the 'what is known' box (and in the conclusion) this should be 'some' probiotics reduce NEC.....
And in the 'what this study adds' the point on NEC should be removed as this data is ND outcome, not NEC

*Thank you. We have added 'some probiotics' as requested, added in 'this probiotic combination' and therefore left the NEC in the manuscript.*

**VERSION 2 – REVIEW**

| | |
|---|---|
| **REVIEWER** | vermeulen, marijn<br>ErasmusMC Rotterdam<br>The Netherlands<br>Competing interests: none |
| **REVIEW RETURNED** | 28-Sep-2017 |

| | |
|---|---|
| **GENERAL COMMENTS** | I am pleased with the adjustments that have been made according to the earlier reviewers suggestions. In my opinion, the revised manuscript is worth publishing in BMJ Peadiatrics in its current form. |

| | |
|---|---|
| **REVIEWER** | Berrington, Janet<br>Newcastle upon Tyne Hospitals NHS Foudnation TRust and Newcastle University, UK<br>Competing interests: None |
| **REVIEW RETURNED** | 29-Sep-2017 |

| | |
|---|---|
| **GENERAL COMMENTS** | The authors have addressed some issues, but for me at least there remains a lack of information about the methods of this specific study (as opposed to the ProPrems itself) and the perspective that it has been written from. It still feels odd to phrase things (abstract |

| | |
|---|---|
| | concusion) as ‘Adminstration of the probiotics combination Bifidobacterium infantis, Streptococcus thermophilus and Bifidobacterium lactis to very preterm babies from soon after birth until discharge home or term CA, did not affect survival free of major disability in early childhood. This probiotics mixture reduced necrotising enterocolitis without adverse effects on neurodevelopment or behaviour in very preterm children’ when the elephant in the room is that we would have expected outcome to be better where we see a reduction in NEC and LOS. As pointed out by reviewer two there are no real reasons to expect adverse effect, and yet the conclusion in the abstract is still this (and many readers will only read the abstract)<br>It still does not carry information on:<br>how parents were approached to consider taking part in the follow up study (mail/phone), by whom (local PI or study CI), how many times they were approached before being considered 'lost' to follow up, why the assessment was targeted at 2.5 years, who actually did the ND assessments (local team or study specific team), whether local data was sought if the study team did not make contact (I assume that local f/u was being done in these infants), what the power of the study was thought to be at the front end, etc. This is all really important as it affects who says yes and is therefore assessed.<br>Likewise the data presented about age at assessment does not really match the stated intent to examine at 2.5 years – we have a mean of 30 months, with an SD around 7 or 8 months, implying that an important number of infants were assessed significantly before their second birthday? If the approach was at 2.5 years how are some assessed so much sooner?? If the intent was to assess at 2.5 years I would have expected the mean (median) to be higher than 2.5 years? We might benefit from more descriptive data for this, such as the IQR’s to help understand when the assessments were actually done, and how this fits with the applied methodology seeking parental attendance.<br>Information pertaining to approvals is now presented, but in the outcomes section not the methods section<br>The power of the study is important to address. I find the table presented confusing as there are no associated p values, but get the point that the non-attending group would have to be very different from the assessed group to change the outcome. I would personally prefer an up front power calculation, simply telling us what the study power was to detect an x% difference between the two groups for the primary outcome given what n was, as I think most readers are used to that and understand what it means. If this study then has 50% power to detect a difference of 10% between groups that is important (it may well be better!) |

## VERSION 2 – AUTHOR RESPONSE

Reviewer: 1

**Comments to the Author**

I am pleased with the adjustments that have been made according to the earlier reviewers suggestions. In my opinion, the revised manuscript is worth publishing in BMJ Peadiatrics in its current form.

Thank you.

Reviewer: 2

**Comments to the Author**

The authors have addressed some issues, but for me at least there remains a lack of information about the methods of this specific study (as opposed to the ProPrems itself) and the perspective that it has been written from. It still feels odd to phrase things (abstract concusion) as 'Adminstration of the probiotics combination Bifidobacterium infantis, Streptococcus thermophilus and Bifidobacterium lactis to very preterm babies from soon after birth until discharge home or term CA, did not affect survival free of major disability in early childhood. This probiotics mixture reduced necrotising enterocolitis without adverse effects on neurodevelopment or behaviour in very preterm children' when the elephant in the room is that we would have expected outcome to be better where we see a reduction in NEC and LOS. As pointed out by reviewer two there are no real reasons to expect adverse effect, and yet the conclusion in the abstract is still this (and many readers will only read the abstract).

As requested by Reviewer 2 and the editors, the Abstract Conclusions have been rephrased as: 'Administration of the probiotics combination Bifidobacterium infantis, Streptococcus thermophilus and Bifidobacterium lactis to very preterm babies from soon after birth until discharge home or term CA, did not adversely affect neurodevelopment or behaviour in early childhood.'

It still does not carry information on:
how parents were approached to consider taking part in the follow up study (mail/phone), by whom (local PI or study CI), how many times they were approached before being considered 'lost' to follow up, why the assessment was targeted at 2.5 years, who actually did the ND assessments (local team or study specific team), whether local data was sought if the study team did not make contact (I assume that local f/u was being done in these infants), what the power of the study was thought to be at the front end, etc. This is all really important as it affects who says yes and is therefore assessed. Likewise the data presented about age at assessment does not really match the stated intent to examine at 2.5 years – we have a mean of 30 months, with an SD around 7 or 8 months, implying that an important number of infants were assessed significantly before their second birthday? If the approach was at 2.5 years how are some assessed so much sooner?? If the intent was to assess at 2.5 years I would have expected the mean (median) to be higher than 2.5 years? We might benefit from more descriptive data for this, such as the IQR's to help understand when the assessments were actually done, and how this fits with the applied methodology seeking parental attendance.
Information pertaining to approvals is now presented, but in the outcomes section not the methods section.

The following has been added to the Assessment subsection of the Methods:
'Our intention was for neurodevelopment and behavioural assessments to be performed at 2 years CA. However, as neurodevelopment was not an outcome of the original ProPrems trial, funding and ethical approval was obtained after some participants were older than 2 years CA. Therefore assessments were performed as close to 2 years CA as possible, and before 5 years CA. Families of surviving ProPrems participants were made aware of the follow-up study in regular study newsletters, and were then approached by mail and then by telephone 3 weeks later about participation in follow-up assessments. Where assessments had already been performed at 2 years CA as part of routine local follow-up, retrospective consent was requested to include the data.'

The power of the study is important to address. I find the table presented confusing as there are no associated p values, but get the point that the non-attending group would have to be very different

from the assessed group to change the outcome. I would personally prefer an up front power calculation, simply telling us what the study power was to detect an x% difference between the two groups for the primary outcome given what n was, as I think most readers are used to that and understand what it means. If this study then has 50% power to detect a difference of 10% between groups that is important (it may well be better!)

The following has been added to the Statistical Analysis subsection of the Methods:
'The ProPrems randomised trial sample size of 1100 participants had 80% power to detect a difference of 7% (16% probiotics vs 23% placebo) in the incidence of culture proven sepsis. Assuming a similar follow-up rate of 90% to that seen in other Victorian cohorts at 2 years CA, this neurodevelopmental follow-up study would have more than 80% power to detect a difference of 8% (82% probiotics, 74% placebo) in the primary outcome (survival free of major neurosensory impairment at 2 years' CA).'

And the following has been added to Paragraph 6 of the Discussion:
'However, with 373 intervention and 362 control participants, there was 75% power to detect a difference of 82% vs 74% in survival free of major neurosensory impairment. Also, those…'