PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	CLINICAL UTILITY OF CORPUS CALLOSUM MEASUREMENTS IN HEAD SONOGRAMS OF PRETERM INFANTS: A COHORT STUDY
AUTHORS	Perenyi, Agnes; Amodio, John; Katz, Joanne; Stefanov, Dimitre

VERSION 1 - REVIEW

REVIEWER	Dr Nigel Anderson Dept of Academic Radiology University of Otago, Christchurch Christchurch New Zealand
REVIEW RETURNED	24-Jan-2013

GENERAL COMMENTS	RESEARCH QUESTION - Study Hypothesis : corpus callosum measurement at routine head ultrasound in premature infants would help early identification of infants at risk of neurodevelopmental delay. STUDY POPULATION Study population was 502 subjects born at 23-36 weeks. 218 were excluded. Of the remaining 284, 173 had neurodevelopmental follow-up. Of these 173, 58 were born at 23-29 weeks. 1. The authors should limit their study to those with neurodevelopmental workup, as that is the goal of their study. METHODS: Corpus callosal measurements were obtained 1-6 times per infant, and correlated with short-term clinical features and neurodevelopmental outcome at 18-22 months of corrected age. 2. Which CC measurements at which post-natal age were correlated with outcome? Was the measurement at birth or discharge or all measurements? The group of 87 born 23-29 weeks are the most interesting sub-group. 3. Was growth rate of CC calculated? 4. How did growth rate of CC correlate with ND outcome? Need

	clearer and more detail about which CC measurements were used for correlation. ABSTRACT and SUMMARY: 5. The results have not been presented in sufficient detail to know if the authors can conclude that CC measurement does not help predict ND outcome (see above - more detail of which CC measurements were used for correlation purposes). 6. The statement that CC measurement has no role in predicting ND outcome is too sweeping based on the limited results provided, and deserves more caveats than just about ethnicity of the study population. RESULTS and CONCLUSION: See above for questions on which CC measurements were used for correlation with ND outcome. 7. The statement that "In agreement with Cooke et al., 19 we found no significant correlation between ND outcome and imaging study results (CC measurements) in this study, although CC measurements had been performed at a different age and by different modality (MRI). " is very brief and hides a lot of differences between the two studies. 8. How do the authors' results compare with Anderson et al articles which have a similar methodology, or Thomson et al articles based on MR of corpus callosum? Thompson DK, Inder TE, et al several articles on corpus callosal length and development in VLBW infants eg Neuroimage. 2011 Mar 15;55(2):479-90 *And* Neuroimage. 2012 Feb 15;59(4):3571-81. COMMENT: White matter loss in VLBW infants is a serious complication. The hypothesis that corpus callosal monitoring might help predict outcome is a valid hypothesis. 8. Why do the authors think the results of their study did not show correlation with neurodevelopmental outcome at 20 months of age? Is it methodology – what about growth rate of corpus callosum?, is the 20 month age at assessment too early? 9. What do the authors suggest is a better way to predict neurodevelopmental outcome?

REVIEWER	Deanne Kim Thompson, PhD Senior Research Officer Murdoch Childrens Research Institute Australia
REVIEW RETURNED	25-Jan-2013

THE STUDY	Research question - Aim one is not well defined or motivated. Overall study design - I have reservations about the exclusions and represenativeness of the cohort used to answer the research questions. Participant description - the inclusion/exclusion criteria, and subset used for analysis, and differences between them are not clearly defined. Patients representative - See above, and in addition only African-American infants are represented. Methods - I would like to have some sections of the methods better described. Statistics - These may well be appropriate, but they seemed a little more complex than they needed to be. References - Some of the literature in the introduction and discussion did not seem relevant.
RESULTS & CONCLUSIONS	I think the message could be clearer, especially once aim 1 is better defined and clarified.
GENERAL COMMENTS	Summary: This manuscript aims to determine 1) whether corpus callosum size, as measured on head sonograms, is related to perinatal morbidities such as bronchopulmonary dysplasia, patent ductus arteriosus, sepsis, or retinopathy of prematurity, and 2) whether corpus callosum size is clinically useful for predicting early childhood adverse neurodevelopmental outcomes. The authors found that CC measurements did not relate to perinatal morbidities, nor did they provide clinical usefulness in predicting adverse outcome. Abstract: Please be consistent with what you call your corpus callosum measure throughout the manuscript – in the abstract objectives you use 'thickness', but elsewhere you use size (length and thickness). You have not given a hypothesis regarding the relationship of corpus callous size and perinatal morbidities. In the interventions section, it would be great if you could say, even briefly, what kind of clinical variables were collected and what kind of neurodevelopmental data you have. Introduction: What is the relevance of corpus callosum growth in the context of this study (in second paragraph of introduction)? Aim 1 is not motivated well in the introduction, and indeed is not very well integrated into the manuscript as a whole. It is not really clear why you are doing this. Why would corpus callosum size be expected to predict lung disease, a patent duct, infection or retinopathy? Presumably the corpus callosum measures are taken before these morbidities are diagnosed? I would have thought it would be possibly that these morbidities predict corpus callosum size, but not the other way around. Patient population: Why were those with abnormalities on head sonograms excluded? I

	am concerned that the study may only be representative of 'healthier' preterm populations, which possibly defeats the purpose of using this method to predict those with poor outcomes. If you have excluded late preterm infants, I wonder if the focus of the study should therefore be on very preterm infants? And in this case, why do you use a cut-off of 34 rather than 32 weeks? The 34 week cut-off seems neither here nor there. Methods: Which test was done at which time point? You mention 2 timepoints: 18-22 months, and up to 3.5 years, but it is not clear what was done when. This seems like a large age range for testing. Please make it clear what the developmental tests (eg. DDST and CAT/CLAMS) are actually testing. It wasn't clear to me in this section (until later) that you are just using one overall measure of adverse neurodevelopmental outcome, so please clarify. Statistical analysis: Please use the word 'outcome' consistently. Here you have used 'outcome' to describe the perinatal morbidities as well as the neurodevelopmental outcomes, which is confusing. What analysis is the subgroup of 87/284 infants used for and why? You have not described how your measure of corpus callosum size was computed – is it a single composite measure of both length and thickness? Or were length and thickness tested separately? It appears that you adjusted for gestational age at follow-up, but what about adjusting for gestational age at head sonogram? I must say, as a non-statistician, I do not really understand your statistical models? Can you justify why you have used the statistical models you use? If you used more simple statistical tests, such as a simple regressions between the corpus callosum size and either morbidities or adverse outcomes, do you draw the same conclusions? Results: I am confused with the two populations you are reporting on – one (in table 1) with n=284 - those selected/eligible under your criteria, then n=173 with follow-up data, but then the sub-set, which it appears all analyses are based on, with n=87? If you are only analyzing n=87, shouldn't these be the ones reported on throughout the manuscript, including in Table 1? Then you would have to report if there were any differences between those analyzed vs. those not analyzed. I wonder if all the text under Table 1 shoud be incorporated into the table. Also, it would be helpful if the types of adverse neurodevelopmental outcomes you have mentioned here were explained in the methods section. I am afraid I don't really understand Figure 1, but it may just be my lack of statistical understanding! Discussion: Is the paragraph about late preterm infants relevant, since you excluded the late preterm infants from your study? In your final summary, I am not sure you can generalize that corpus callosum head sonograms do not have any clinical utility in predicting adverse neurodevelopmental outcome in preterm infants since you had so many exclusions.

VERSION 1 – AUTHOR RESPONSE

Reviewer: Dr Nigel Anderson
Dept of Academic Radiology
University of Otago, Christchurch
Christchurch
New Zealand

STUDY POPULATION:
1. The authors should limit their study to those with neurodevelopmental workup, as that is the goal of their study.

We have eliminated patients from the study who did not have neurodevelopmental follow-up.

METHODS:
2. Which CC measurements at which postnatal age were correlated with outcome? Was the measurement at birth or discharge or all measurements?
The group of 87 born 23-29 weeks are the most interesting sub-group.

We included only the 87 infants who had head sonograms between 23 and 29 weeks.

3. Was growth rate of CC calculated?

We did not calculate growth rate of the CC. Sixty-one (70%) of the infants either had only one HUS, or only one that had optimal quality for obtaining CC measurements during the data collection time frame of 0-6 postnatal weeks.

4. How did growth rate of CC correlate with ND outcome? Need clearer and more detail about which CC measurements were used for correlation.

Since we did not calculate the growth rate of the CC, we could not correlate it with ND outcome. In the method section, we further clarified which CC measurements have been used.

ABSTRACT and SUMMARY:
5. The results have not been presented in sufficient detail to know if the authors can conclude that CC measurement does not help predict ND outcome (see above - more detail of which CC measurements were used for correlation purposes).

Please see our reply on number 4.

6. The statement that CC measurement has no role in predicting ND outcome is too sweeping based on the limited results provided, and deserves more caveats than just about ethnicity of the study population.

In our summary, we have concluded that the first measurements of the CC are not predictive of short-term ND outcome (i.e. outcomes between 18-22 months corrected age).

RESULTS and CONCLUSION:
7. The statement that "In agreement with Cooke et al., 19 we found no significant correlation between ND outcome and imaging study results (CC measurements) in this study, although CC measurements had been performed at a different age and by different modality (MRI). " is very brief and hides a lot of

differences between the two studies.

In our revised manuscript, we further clarify the comparison of our study that of with Cooke et al.

8. How do the authors' results compare with Anderson et al articles which have a similar methodology, or Thomson et al articles based on MR of corpus callosum?

The articles by Anderson et al (2005, 2006) primarily describe growth rate and impaired growth of the CC in preterm infants. Since our study does not address growth rate of the CC and we measured and analyzed both the length and thickness of the CC, comparison of our study with those of Anderson et al. is limited.

The articles by Thompson et al (2011, 2012) examine the CC in preterm infants using MRI. Since we used head sonograms rather than MRI in our methodology, our comparison with Thompson et al is also limited. Furthermore, the types of measurements in these studies involved measurement of surface area, cross-sectional area and shape analysis, rather than measurement of length and thickness, as we measured in our study.

COMMENT:
8. Why do the authors think the results of their study did not show correlation with neurodevelopmental outcome at 20 months of age? Is it methodology – what about growth rate of corpus callosum?, is the 20 month age at assessment too early?

We speculate that measuring two dimensions (length and thickness) of the CC may not have as good predictive value the techniques described in the studies above. It is possible that the measurement of CC thickness may have less value because its reproducibility is not as robust as that of length measurement.

9. What do the authors suggest is a better way to predict neurodevelopmental outcome?

We may use growth assessment of the CC as Anderson et al reported. Although brain MRI as a method may be more accurate in general, head sonograms are easier to administer to patients as they may be done at bedside, do not require sedation and are more cost-effective.

Although it is outside the scope of our study, we feel that analysis of infant general movements according to Prechtl is a better way to predict ND outcome. (Einspieler C et al, Prechtl's Method on the Qualitative Assessment of General Movements in Preterm, Term and Young Infants, Mac Keith Press, London, 2004) This assessment has been incorporated in our standard of care evaluating ND outcome.

Reviewer: Deanne Kim Thompson, PhD
Senior Research Officer
Murdoch Childrens Research Institute
Australia

Research question - Aim one is not well defined or motivated.

We have better clarified the objectives of the study.

Overall study design - I have reservations about the exclusions and represenativeness of the cohort used to answer the research questions.

The cohort of infants included in the study is representative of our patient population. Since the inclusion/exclusion criteria are mentioned several times by this reviewer, we will address it below.

Participant description - the inclusion/exclusion criteria, and subset used for analysis, and differences between them are not clearly defined.

Exclusion criteria are based upon patient issues or issues regarding sonogram technique.

a) Infants who died were excluded because either they did not have a head sonogram or they never had any follow-up.

b) We excluded the 10 infants of non-African-American origin. Since the number is small, it would not have made a significant different in the data analysis. Our hospital is located in an urban community where the population is 98% African-American.

c) In the revised manuscript, we excluded those infants who survived but did not have subsequent ND follow-up. Since the most immature infants have the highest risk of adverse ND outcome, we focused on the 87 infants who had their first sonograms between 23 and 29 weeks (the first weeks of life).

d) Based on the sonogram technique, infants with head sonograms of suboptimal quality were excluded because the CC could not be clearly visualized and measured. We excluded infants with brain development anomalies, intracranial bleeding and periventricular leukomalacia for two reasons: first, these changes in the brain imaging study may make CC measurements difficult or inaccurate. Second, all of these pathologies carry their own risk for adverse ND outcome.

Methods - I would like to have some sections of the methods better described.

In the revised manuscript, we have improved the descriptions in the methods section.

Statistics - These may well be appropriate, but they seemed a little more complex than they needed to be.

We have revised the statistical methods sections. Please see our responses to the questions below related to statistical analysis.

References - Some of the literature in the introduction and discussion did not seem relevant.

We removed irrelevant references from the revised manuscript.

I think the message could be clearer, especially once aim 1 is better defined and clarified.

We improved the clarity of the key messages.

Abstract:
Please be consistent with what you call your corpus callosum measure throughout the manuscript – in the abstract objectives you use ‘thickness’, but elsewhere you use size (length and thickness).

In the revised manuscript, we have used descriptions of CC measurements consistently.

You have not given a hypothesis regarding the relationship of corpus callous size and perinatal morbidities.

In the revised manuscript in the discussion we describe the possible association between morbidities related to prematurity and CC measurements.

In the interventions section, it would be great if you could say, even briefly, what kind of clinical variables were collected and what kind of neurodevelopmental data you have.

Table 1 contains a description of the clinical variables of the cohort of infants in our study. We also indicate the types of adverse ND outcome that were diagnosed.

Introduction:
What is the relevance of corpus callosum growth in the context of this study (in second paragraph of introduction)?

CC size (length and thickness) is related to CC growth. We have revised the language in this section of the introduction.

Aim 1 is not motivated well in the introduction, and indeed is not very well integrated into the manuscript as a whole. It is not really clear why you are doing this.

We have better clarified the objectives of the study.

Why would corpus callosum size be expected to predict lung disease, a patent duct, infection or retinopathy? Presumably the corpus callosum measures are taken before these morbidities are diagnosed? I would have thought it would be possibly that these morbidities predict corpus callosum size, but not the other way around.

In the last paragraph of the introduction, we wrote that we hypothesized that CC measurements in head ultrasounds may be clinically useful to predict ND outcome. We did not indicate that CC size would be expected to predict morbidities.

Patient population:
Why were those with abnormalities on head sonograms excluded? I am concerned that the study may only be representative of 'healthier' preterm populations, which possibly defeats the purpose of using this method to predict those with poor outcomes.

We excluded infants with brain development anomalies, intracranial bleeding and periventricular leukomalacia for two reasons: first, these changes in the brain imaging study may make CC measurements difficult or inaccurate. Second, all of these pathologies carry their own risk for adverse ND outcome.

The subset of the population that is included in the revised manuscript includes those infants with head ultrasound (HUS) studies between 23-29 weeks. This group of infants is not representative of a 'healthier' preterm population, as 32% were diagnosed with BPD, 31% with PDA, 24% with ROP, 48% with sepsis, and 64% with composite morbidity.

If you have excluded late preterm infants, I wonder if the focus of the study should therefore be on very preterm infants? And in this case, why do you use a cut-off of 34 rather than 32 weeks? The 34 week cut-off seems neither here nor there.

In the revised manuscript, every infants' gestational age is under 32 weeks.

Methods:

Which test was done at which time point? You mention 2 timepoints: 18-22 months, and up to 3.5 years, but it is not clear what was done when. This seems like a large age range for testing.

In our clinic, we follow infants up to 3.5 years of age. In the study, each infant had ND evaluation between 18-22 months of age. We have revised this section of the manuscript for improved clarity.

Please make it clear what the developmental tests (eg. DDST and CAT/CLAMS) are actually testing. It wasn't clear to me in this section (until later) that you are just using one overall measure of adverse neurodevelopmental outcome, so please clarify.

We have revised the manuscript to improve the clarity of this issue.

Statistical analysis:
Please use the word 'outcome' consistently. Here you have used 'outcome' to describe the perinatal morbidities as well as the neurodevelopmental outcomes, which is confusing.

Although morbidities related to prematurity are considered to be outcomes in the clinical setting, we understand that these diagnoses are not outcomes within the context of the study.

What analysis is the subgroup of 87/284 infants used for and why?

We focused on the infants between 23-29 weeks in order to see if CC measurements during this timeframe (0-6 weeks postnatal age) would be predictive of ND outcome. This group included the 87 infants whose data were analyzed.

You have not described how your measure of corpus callosum size was computed – is it a single composite measure of both length and thickness? Or were length and thickness tested separately?

CC length and thickness were measured separately but tested together. We have made this procedure more evident in the revised manuscript.

It appears that you adjusted for gestational age at follow-up, but what about adjusting for gestational age at head sonogram?

Thank you for this suggestion. In the new submission we adjusted for postnatal age of the CC measurements. Statistically, this is equivalent for adjusting for the gestational age at the time of the head sonograms, since we also adjusted for gestational age at birth.

I must say, as a non-statistician, I do not really understand your statistical models? Can you justify why you have used the statistical models you use? If you used more simple statistical tests, such as a simple regressions between the corpus callosum size and either morbidities or adverse outcomes, do you draw the same conclusions?

We revised the statistical section, explaining in detail the rationale for using the chosen statistical analysis. We used logistic regression since the outcome (ND outcome) is binary. Linear regression, for comparison, is an appropriate method used for continuous outcome variables. The individual p-values for CC length and thickness from the logistic regression model answers the questions if any of the CC measurements (or both of them jointly, which is based on a separate test with df=2) are significantly associated with the outcome, adjusted for the other covariates. This is an important question, but the more specific question, which addresses our hypothesis, is whether the CC measurements provide additional predictive power beyond a set of easily available clinical variables. We used the area under the ROC curve (AUC), since it is the most widely used measure of predictive

performance. Thus, our initial hypothesis can be addressed by determining whether the improvement of the AUC after including CC measurements was statistically significant. We used the method of DeLong et al. since it is specifically designed to answer this question. All statistical comparisons mentioned above are based on the logistic regression model.

We do not agree with the earlier statement that our statistical methods are more complex than they need to be. To further illustrate that our methods are not very complex, we would like to mention that all calculations (including the graph in Fig. 1) are directly incorporated in the statistical software (we used PROC LOGISTIC in SAS, version 9.2).

Results:
I am confused with the two populations you are reporting on – one (in table 1) with n=284 - those selected/eligible under your criteria, then n=173 with follow-up data, but then the sub-set, which it appears all analyses are based on, with n=87? If you are only analyzing n=87, shouldn't these be the ones reported on throughout the manuscript, including in Table 1? Then you would have to report if there were any differences between those analyzed vs. those not analyzed.
I wonder if all the text under Table 1 should be incorporated into the table. Also, it would be helpful if the types of adverse neurodevelopmental outcomes you have mentioned here were explained in the methods section.

Thank you for the suggestion. The revised Table 1 simply describes clinical variables of the subset of 87 infants whose CC measurements had been analyzed.

We do not think that description of adverse ND outcome can easily be incorporated into Table 1. The types of ND outcome that we found do not belong in the method section, as this is part of our results.

I am afraid I don't really understand Figure 1, but it may just be my lack of statistical understanding!

We presented the two ROC curves in Fig. 1, for the models with and without CC measurements, as ROC curves are commonly presented. ROC curves are informative, beyond the information captured by the AUC. We hope that providing more detail in the statistical section makes Fig. 1 relevant.

Discussion:
Is the paragraph about late preterm infants relevant, since you excluded the late preterm infants from your study?

We removed the paragraph regarding late preterm infants.

In your final summary, I am not sure you can generalize that corpus callosum head sonograms do not have any clinical utility in predicting adverse neurodevelopmental outcome in preterm infants since you had so many exclusions.

In our final summary, we revised our conclusions based upon the inclusion criteria of the cohort, and gave our opinion regarding the clinical usefulness of early CC measurements and other methods for predicting ND outcome.

VERSION 2 – REVIEW

REVIEWER	Dr Nigel Anderson Department of Academic Radiology and Centre for Bioengineering, University of Otago, Christchurch 2 Riccarton Avenue Christchurch New Zealand
REVIEW RETURNED	21-Mar-2013

GENERAL COMMENTS	The authors have improved the manuscript and answered my concerns. My only remaining quibble is that "short-term" should be added wherever "neurodevelopmental outcome" is written. It is added in many but not all instances. It particularly needs to be in the "article focus" section. I do not think further review of such a minor change is required, except by the editor.

REVIEWER	Deanne Kim Thompson, PhD Senior Research Officer Murdoch Childrens Research Institute Australia
REVIEW RETURNED	15-Mar-2013

THE STUDY	The aims/hypotheses are a lot better but could still be clarified and motivated better
GENERAL COMMENTS	This manuscript is much clearer! I have some fairly picky suggestions which I hope will make it even clearer - feel free to change these around a bit if they don't make sense! Abstract: At the end of your hypothesis in the objective section, I would add '..., over and above the predictive power of perinatal morbidities.' In the interventions section, I would add 'CC size (length and thickness) was measured...' And at the end of the last sentence on statistics, I would add '..., while adjusting for perinatal morbidities.' In results section, I would say 'Measurements of CC size did not add substantial power to predict short-term ND outcome beyond the information provided by the presence of morbidities (retinopathy of prematurity, bronchopulmonary dysplasia, patent ductus arteriosus and sepsis). In the conclusion, I would delete the first sentence. Introduction: I would rework the second paragraph so that the sentence 'Anderson et al...' goes after the first sentence, and then link the next sentence, i.e. 'Thus, premature infants also....' I also think the last paragraph needs to be reworked, maybe 'Morbidities associated with prematurity, such as retinopathy of prematurity, bronchopulmonary dysplasia, patent ductus arteriosus and sepsis, may also influence brain development, and are therefore often used to predict ND outcome (reference). This study aimed to determine if CC size measurements from routine HUS may be clinically useful to predict short-term ND outcome, over and above the predictive power of perinatal morbidities.

	Patient population: For the final 2 sentences where you are describing that there are only 87 infants, it may be clearer to put it like this: 'For the purposes of this study, only those with early HUS (between 23-29 weeks' gestational age or 0-6 postnatal weeks) were analyzed, being 87 of the remaining 173 infants.' Methods: How many cases were repeated for intra and inter-rater reliability? Statistical analysis: I would delete the first sentence as you have already explained this in the previous section. I would reword the second paragraph as follows: 'Logistic regression was used to determine if CC length and thickness were associated with ND outcome when adjusted for four perinatal morbidities (ROP, PDA, BPD, and sepsis). In addition, all analyses were also adjusted for GA, postnatal time of HUS and gender.' - I don't think you have defined ROP, PDA, BPD and so should spell them out in full? Results: What are the numbers for the infants included or excluded, in paragraph 2? I.e. put 'included (n=87)...', 'excluded (n=) with regard...' I wonder if you should report the prediction of CC size on ND outcome both BEFORE and after adjusting for the other perinatal variables? Were there associations before adjusting? Discussion: In the discussion, the points made in relation to previous studies are not well linked to the current study, for example in the first paragraph, you could begin with 'Contrary to the current findings...' I am not sure you need the second sentence 'The size...' You probably don't need to talk about the first Thompson study characterizing the CC, maybe just the second one associating with ND outcome, and the Cooke study, and make it clearer earlier that you are talking about these 2 studies because they are in agreement with the current study. Then delete 'This finding is in agreement with that of...' from the 3rd last paragraph.

VERSION 2 – AUTHOR RESPONSE

Reviewer: Dr Nigel Anderson
Dept of Academic Radiology
University of Otago, Christchurch
Christchurch
New Zealand

Thank you for your repeat review. Following your advice, we added "short-term" wherever we mentioned ND (neurodevelopmental) outcome when it was appropriate; if the quoted publications include long-term outcome (i.e. longer than 18-22 months of corrected age) or the outcome was spelled out (like "outcome at 2 years of age"), we omitted the "short-term" before the "ND outcome."

Reviewer: Deanne Kim Thompson, PhD
Senior Research Officer
Murdoch Childrens Research Institute

Australia

Thank you for your repeat review and suggestions.
In response, please note the following

-Abstract: we added the improved wording according to your suggestion 1. at the end of the hypothesis (…over and above)
- we added to the last sentence in the 'Statistics' ("…while adjusting)
- we changed the "Result" section and deleted the first sentence.

-Introduction: we reworked the second paragraph and the last paragraph according to your suggestion

-Patient population: -we rephrased the final two sentences according to your suggestion

-Methods: - we provided and included information regarding the intra- and interrater reliability.

-Statistical analysis: - we deleted the first sentence of this section.
-we reworded the second paragraph according to your suggestion

- we spelled out the abbreviations that stand for morbidities related to prematurity (including bronchopulmonary dysplasia [BPD], retinopathy of prematurity [ROP], patent ductus arteriosus [PDA]) in the third paragraph of the Introduction. We provided additional information regarding these morbidities in the fourth paragraph of the Methods.

-Results: - your comment on this was not clear to us, so we did not change this section. We did not analyze ND outcome before adjusting for variables.

-Discussion: - we reformulated the first paragraph and deleted the second sentence.
- we deleted the first Thompson study.
- we deleted the sentence … "This finding…

Thank you for your attention.