

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)	Physiological Stress Responses in Infants at 29-32 weeks' Postmenstrual Age During Clustered Nursing Cares and Standardized Neurobehavioural Assessments
AUTHORS	Allinson, Leesa; Denehy, Linda; Doyle, Lex; Eeles, Abbey; Dawson, Jennifer; Lee, Katherine; Spittle, Alicia

VERSION 1 - REVIEW

REVIEWER	Pineda, Bobbi Washington University, Occupational Therapy, Pediatrics Competing Interest: None
REVIEW RETURNED	16-Mar-2017

GENERAL COMMENTS	<p>This study investigated the differences in heart rate and oxygen saturation levels and events in very preterm infants, being assessed early, during neurobehavioral testing and routine nursing cares. Very important contribution to the literature, as it is important to understand the impact of neurobehavioral testing in early gestation during NICU stay.</p> <p>I would suggest the following clarifications: in the section about heart rate and oxygen saturation, define exactly what was captured from the video. Were there time points that were assessed for the numeric value and then averaged. The use of thresholds like brady and tachy are more obvious but the changes in these measures needs some clarification. In addition, why 180 as the threshold for tachy (US usually uses 200). How did the researchers account for the fact that the pulse oximeter is sensitive to movement and may not have registered correctly during the neurobehavioral assessment. How did the neurobehavioral test compare to baseline readings (ie is it possible that the fluctuations are similar when the infant is and is not being tested)?</p>
-------------------------	--

REVIEWER	Danks, Marcella Australian Catholic University, School of Physiotherapy Competing Interest: I have no conflict of interest. I have seen this work in another format previously as I reviewed the first author's PhD submission.
REVIEW RETURNED	17-Mar-2017

GENERAL COMMENTS	This research was well conducted and is well presented with valid
-------------------------	---

	<p>interpretation of findings. I have a few queries which are outlined in the review below.</p> <p>Title: Makes a clear statement of what research was undertaken.</p> <ul style="list-style-type: none"> • Abstract: provides an informative and balanced summary of what was done and what was found • Introduction: provided clear rationale for the study. There is a clear statement of Aims 1&2. <p>Methods:</p> <ul style="list-style-type: none"> • Study design clearly stated. • Setting described. • Participants defined with clear inclusion and exclusion criteria • Assessments outlined clearly. The reference for these assessments is an excellent review article however I recommend the authors use the primary reference when first referring to these assessments. • Protocol including procedures to protect standardization of measures and blinding clearly outlined. • Calculation of sample size to achieve adequate power in the study is reported. Statistical analysis is justified. <p>Results:</p> <ul style="list-style-type: none"> • Clearly reported generally. • Please be more specific in your terminology regarding 'lower HR' n p 7, last paragraph, line 4: do you mean lower HR as in not tachycardic or lower HR as in bradycardic? • Please also provide more details about the rare bradycardic events. When did they occur? During which assessment? • Please check the reported values for mean difference p8 paragraph 2. • Please make further comment regarding reported desaturations during neurobehavioural assessment: Were they episodes that were considered clinically significant? Was there any significant difference in episodes of desaturation between clustered nursing care and neurobehavioural assessment (whether it required handling or not). <p>Discussion and Conclusion:</p> <ul style="list-style-type: none"> • Key results summarized. • It is interesting to note your report that the neurobehavioural assessment that does not require handling has a greater risk of episodes of desaturation than the assessments that do require handling. Please comment regarding the clinical importance of the risk of desaturation with neurobehavioural assessment which does or does not require handling. • P 11 paragraph 2, last line: Please note which of the statistically significant results you considered not to be clinically important. <p>References: please check reference 2.</p>
--	--

VERSION 1 – AUTHOR RESPONSE

Reviewer 1 Comments:

In the section about heart rate and oxygen saturation, define exactly what was captured from the video.

Author Response: In the Methods at the bottom of page 6 we state “HR, SpO2, plethysmograph wave, signal identification and quality indicator (signal IQ) and alarm message data were extracted at five-second intervals during the assessment/care by AE, and entered into a Microsoft Excel spreadsheet by LA.

Were there time points that were assessed for the numeric value and then averaged?

Author Response: No – the data recorded at each 5- second interval were analysed using linear regression, fitted using generalized estimating equations (GEEs) to allowing for multiple observations within individual participants, as stated in the Methods under “Statistical Analysis” on page 7.

The use of thresholds like brady and tachy are more obvious but the changes in these measures needs some clarification.

Author Response: The frequencies of episodes of bradycardia and tachycardia are provided in Table 2.

Why 180 as the threshold for tachy (US usually uses 200)?

Author Response: In Australia, >180bpm is the standard threshold for tachycardia in infants supported by The WHO recommended guidelines for Integrated Management of Pregnancy and Childbirth: Pregnancy, Childbirth, Postpartum and Newborn Care; A guide for essential practice, 2003 and Approach to Pediatric Tachycardia by Justin Haba from the Department of Pediatrics, University of British Columbia. Both references have been added to the manuscript where the tachycardia threshold has first been mentioned.

How did the researchers account for the fact that the pulse oximeter is sensitive to movement and may not have registered correctly during the neurobehavioural assessment?

Author Response: This is stated in the sentence: "During data extraction, motion artefact was determined visually by viewing the pulse amplitude indicator or signal IQ. Only data with good plethysmograph wave and good signal quality with no alarm message (low signal IQ, low perfusion, sensor off, or ambient light) were included in the analysis." Bottom of page 6 and top of page 7.

How did the neurobehavioural test compare to baseline readings (i.e. is it possible that the fluctuations are similar when the infant is and is not being tested)?

Author Response: As our intention was to compare physiological stability between different standards of clinical care, we did not take baseline readings, so we cannot compare our results to baseline, or at rest, values.

Reviewer 2 Comments:

Recommend the authors use the primary reference when first referring to the assessments.

Author Response: The primary references for the assessments have been added.

Please be more specific in your terminology regarding 'lower HR' n p 7, last paragraph, line 4: do you mean lower HR as in not tachycardic or lower HR as in bradycardic?

Author Response: The word “mean” has been added between “lower” and “HR” to indicate that it was the mean HR that was lower during neurobehavioural assessments compared with standard nursing cares.

Please provide more details about the rare bradycardic events. When did they occur? During which assessment?

Author Response: There were three brief episodes of bradycardia during nursing cares - One when an NGT tube was being changed, and two when a baby had its temperature taken. The one bradycardic episode for assessments occurred during GMs; these occurrences have been added on page 12.

Please check the reported values for mean difference p8 paragraph 2.

Author Response: Reported values for means on page 8, paragraph 2 of the Results are correct; we did not report any mean differences in this paragraph. We have added “the mean” before

“postmenstrual age” to avoid any further confusion about what was being reported in this paragraph

Please make further comment regarding reported desaturations during neurobehavioural assessment: Were they episodes that were considered clinically significant? Was there any significant difference in episodes of desaturation between clustered nursing care and neurobehavioural assessment (whether it required handling or not)?

Author Response: There were no serious desaturation episodes requiring resuscitation with either clustered nursing cares or standardised neurobehavioral assessments; this has been added to page 10.

Please comment regarding the clinical importance of the risk of desaturation with neurobehavioural assessment which does or does not require handling.

Author Response: The risk of clinically important desaturations with neurobehavioural assessments is low, and is less frequent than occurs during standard nursing cares, as shown in Table 2.

P 11 paragraph 2, last line: Please note which of the statistically significant results you considered not to be clinically important.

Author Response: We have slightly altered this sentence to “Regardless, the size of the mean difference in SpO₂ of 0.4% was trivial clinically” to indicate that although the mean difference was statistically significant, we did not consider it to be clinically important. We will let readers interpret for themselves if any of the other statistically significant differences is clinically important to them.

Please check reference 2.

Author Response: Reference 2 re-entered.

VERSION 2 – REVIEW

REVIEWER	Koutoumanou, Eirini UCL, UK Competing interests: None
REVIEW RETURNED	11-Oct-2017

GENERAL COMMENTS	<p>This is a well written report with good structure. I am however concerned about one specific aspect of the study design and I would appreciate it if the authors could provide further details and thoughts on this.</p> <p>The authors mention that “clustered nursing cares usually took precedence over the standardised neurobehavioral assessments; the bedside nurse determined the order”.</p> <ul style="list-style-type: none"> - Couldn't the stress caused by the clustered nursing care be carrying an effect on the stress caused by the neurobehavioral assessments (and vice versa for few cases)? - If there is no way to determine the carry over effect of one treatment to the other, the data might be producing distorted results - I appreciate that the authors include this aspect of the data collection in the limitation of the study, however, this still does not make it ok to analyse as is. - For me, this is a major sticking point and I am having difficulty to see what can be salvaged from the given design - (Also, based on what criteria was the nurse allowed to decide the order of ‘assessments’ and how long it would normal be between assessments?) <p>This is an observation study that cannot ascribe causality, hence the</p>
-------------------------	---

	<p>following phrase “standardised neurobehavioral assessments produced less...” should be changed to “standardised neurobehavioral assessments is associated with less...”</p> <p>Some additional comments:</p> <ul style="list-style-type: none"> - Is something missing from the following sentence: “Infants with congenital abnormalities known to affect neurodevelopment” ? - Is the Royal Women's Hospital in Melbourne similar to other hospitals in the area/Australia? - What does PMA stand for in Table 1? - Gestational age is not normally distributed; this was identified by using the following calculation: $(\text{mean} \pm 1.96 * \text{SD})$ which leads to a maximum reference range (30.3) higher than the observed maximum (29). Hence median and inter-quartile range should be reported. Please check for PMA and birthweight as well. - Please include titles for both graphs – I could not locate those but apologies if I am wrong. - What do the whiskers represent in Figure 2 (these are not standard between statistical packages, therefore a clarification is required)? But, I would recommend that you consider swapping the box plots with dot plots that will show all the available data per group and could be accompanied by the mean/median and 95% confidence interval for each group - The unadjusted ORs presented on Table 2 do not seem to match the calculations using the numbers under the Assessment and Cares columns. Can the authors please confirm these are indeed unadjusted ORs? - I would recommend that the following sentence is removed as it is not fully justified: “Given that we had multiple observations for each assessment type for each individual, our study would have more than 81% power to identify such a difference.”
--	--

VERSION 2 – AUTHOR RESPONSE

Reviewer 1 Comments:

The authors mention that “clustered nursing cares usually took precedence over the standardised neurobehavioral assessments; the bedside nurse determined the order”.

- Couldn't the stress caused by the clustered nursing care be carrying an effect on the stress caused by the neurobehavioral assessments (and vice versa for few cases)?
- If there is no way to determine the carry over effect of one treatment to the other, the data might be producing distorted results
- I appreciate that the authors include this aspect of the data collection in the limitation of the study, however, this still does not make it ok to analyse as is.
- For me, this is a major sticking point and I am having difficulty to see what can be salvaged from the given design

Author Response: We agree that clustered nursing cares could cause a carryover effect such that anything coming after such cares, including doing nothing, might be associated with more physiological stress and instability than if it was not after the nursing cares, even if it had no effect on physiological values at all. However, within this study we tried to ensure that the neurobehavioural assessments did not cause more physiological stress and instability than standard nursing cares, which are an essential part of care. Given this, we did not think it was necessary or appropriate to randomly allocate the order of standard cares versus neurobehavioural assessments. We have, however, adjusted for order of assessments in our analysis, which did not change any conclusions – this has been added to the Methods (page 8. Para 1), and to the Results (page 9, last para, page 10, para 1, page 10 para 2, and Table 2 - footnote). An important finding of our study was that most

neurobehavioural assessments occurred after standard nursing cares yet were still associated with less physiological stress and instability than standard nursing cares and therefore makes it even more likely that neurobehavioural assessments are less physiologically stressful than are standard nursing cares.

- (Also, based on what criteria was the nurse allowed to decide the order of 'assessments' and how long it would normal be between assessments?)

Author Response: We allowed the bedside nurse to determine the order without any guidance from the study team in terms of how to make this decision because they know the baby better than the person doing the neurobehavioural assessment and it is reflective of everyday clinical care where the nurse would determine the order of assessments. The following has been added to the methods, page 6, first paragraph to clarify this:

“Clustered nursing cares usually took precedence over the standardised neurobehavioural assessments, as they are an essential component of care. The bedside nurse determined the order based upon the infant’s needs and their workload.”

This is an observation study that cannot ascribe causality, hence the following phrase “standardised neurobehavioral assessments produced less...” should be changed to “standardised neurobehavioral assessments is associated with less...”

Author Response: The statement in the section "What this study adds?" p.3 has been amended to comply with the reviewer’s request, to: “Standardised neurobehavioural assessments are associated with less...”

We have also altered the wording in the Conclusions of the Abstract, and the concluding paragraph of the Discussion in a similar manner.

Is something missing from the following sentence: “Infants with congenital abnormalities known to affect neurodevelopment” ?

Author Response: We have rephrased this sentence to make it clear that this is referring to exclusion criteria

“Infants were excluded if they had a congenital abnormalities known to affect neurodevelopment, had non-English speaking parents, or who were medically unstable“

Is the Royal Women's Hospital in Melbourne similar to other hospitals in the area/Australia?

Author Response: The Royal Women's Hospital in Melbourne is similar to other hospitals in the state of Victoria, and other areas of Australia who have neonatal intensive care units. The following has been added

“from the neonatal nurseries at the Royal Women’s Hospital in Melbourne, a tertiary level neonatal intensive and special care unit”.

What does PMA stand for in Table 1?

Author Response: PMA stands for postmenstrual age. This definition has been added to the footnote at the bottom of Table 1.

Gestational age is not normally distributed; this was identified by using the following calculation: (mean± 1.96*SD) which leads to a maximum reference range (30.3) higher than the observed maximum (29). Hence median and inter-quartile range should be reported. Please check for PMA and birth-weight as well.

Author Response: We agree that gestational age is not normally distributed and neither are PMA nor birthweight, hence we have added median and interquartile ranges to Table 1 for these 3 variables.

Please include titles for both graphs – I could not locate those but apologies if I am wrong.

Author Response: The titles have been added for both graphs.

What do the whiskers represent in Figure 2 (these are not standard between statistical packages, therefore a clarification is required)? But, I would recommend that you consider swapping the box plots with dot plots that will show all the available data per group and could be accompanied by the mean/median and 95% confidence interval for each group.

Author Response: The whiskers in Figure 2 represent the range of the data, up to 1.5 times the interquartile range. An explanation of the whiskers has now been added to Figure 2. The values for median, interquartile range, and range of the data have been calculated from all observations, which comprise multiple values from individuals and hence a dot plot would comprise innumerable dots from the 34 individuals in the study, which would be unhelpful. We would prefer to retain the original box plots as what we are attempting to do is summarise the spread of the data between groups.

The unadjusted ORs presented on Table 2 do not seem to match the calculations using the numbers under the Assessment and Cares columns. Can the authors please confirm these are indeed unadjusted ORs?

Author Response: The unadjusted odds ratios cannot be simply computed from the marginal totals in Table 2 because they have been calculated allowing for the clustering of observations within an individual using GEEs, as explained in the methods. Hence the ORs and their 95% CIs will differ from values calculated just from marginal totals – in particular, the 95% CIs will be typically wider.

I would recommend that the following sentence is removed as it is not fully justified: “Given that we had multiple observations for each assessment type for each individual, our study would have more than 81% power to identify such a difference.”

Author Response: The sentence has been removed as requested.

In addition to the above changes, we have detected a few other typos, most notably that we were inconsistent with the spelling of “neurobehaviour” - it has been changed to English spelling (neurobehaviour) throughout.