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 Archives of Disease in Childhood 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) | Early surfactant and non-invasive ventilation versus intubation and surfactant – a propensity score-matched national study |
|---------------------|-------------------------------------------------------------------------------------------------------------------|
| AUTHORS             | Reigstad, Hallvard  
                        Hufthammer, Karl Ove  
                        Rønnestad, Arild E  
                        Klingenberg, Claus  
                        Stensvold, Hans Jorgen  
                        Markestad, Trond |

VERSION 1 – REVIEW

| REVIEWER             | Reviewer name: Dr. Angela Kribs  
                        Institution and Country: University Hospital of Cologne, Cologne Neonatology and Pediatric Intensive Care, Germany  
                        Competing interests: None |
|----------------------|-------------------------------------------------------------------------------------------------------------------|
| REVIEW RETURNED      | 24-Mar-2022 |

| GENERAL COMMENTS | During the last years less invasive surfactant administration has become more common especially in Europe. Based on some randomised clinical trials and meta-analysis it has been recommended as the preferred mode of surfactant administration in the 2019 European Consensus Guideline. This recommendation was considered weak because some of the underlying studies were open to bias.  
                        Furthermore each new method that has been evaluated in trials conducted under standardized and optimized conditions has also to be evaluated under real live conditions. There may be some concerns that become not obvious under standardized conditions but that are important in everyday practice. For example the experience in performing certain skills may influence safety as well as efficiency of a therapeutic method or a method may be beneficial for a certain group of patients but perhaps even harmful in another population.  
                        The presented paper by Reigstad and coworkers devotes to the comparison of the outcome after LISA compared to primary endotracheal intubation in infants with a gestational age of 25-27 weeks. They performed an analysis of a defined population-based data set of a national registry (Norwegian Neonatal Network). These data represent the results from a real world scenario. The authors compared LISA infants to all non-LISA infants and to those non-LISA infants who received surfactant. In addition they used propensity score matching to identify non LISA infants who were similar regarding potential confounders.  
                        Altogether they found a reduced rate and duration of mechanical ventilation in the LISA infants but no other clear benefits or harms. The paper is important because it presents data of a real world scenario. The method of propensity score matching enables some interesting additional results that are worth to be presented and to be discussed. |


However there are some aspects that need revision.

Introduction
P4, line 42

"It is also a concern that the safe management of LISA requires certain skills, for instance to avoid extended hypoxia during the procedure..."

The same could be mentioned for endotracheal intubation. Each method in the hand of an unexperienced person may result in problems. But if a new method has the potential to improve outcome the lack of experience should not be a reason to hinder its introduction. Then it is the task to define each step of the method, to teach it and thereby to increase experience.

I propose to discuss the differences in experience with endotracheal intubation on the one hand and LISA on the other hand in different centers as one factor influencing outcome, but not as an advantage or disadvantage of the one or the other method.

Discussion
P8, line 35

It is not correct that only the NINSAPP study (Kribs et al 2015) included infants comparable to those analysed in the presented paper. Also the AMV trial (Göpel et al 2011) included infants with a gestational age of 26 and 27 weeks.

P8, line 39

The references (15-24) are obviously not all correct for the statement. Reference 16-19 refer to studies of the INSURE but not of the LISA procedure.

Conclusion

The authors state that individual NICU routines may be equally or more important than LISA vs. non-LISA strategy. This may be the case. But it is also possible that under certain defined NICU routines LISA may be beneficial. Therefore further research should focus on the question which mode of surfactant administration is best under which condition.

I think this point should be discussed and perhaps also be mentioned in the conclusion.

Some minor points

Abstract
P3 line 22-23

"and bronchopulmonary dysplasia at postmenstrual age 36 and 40 weeks"

As bronchopulmonary dysplasia is defined for a postmenstrual age of 36 weeks it would be better to write: "... need of supplemental oxygen or respiratory support at postmenstrual age..."

Material and Methods
P5 line 28 "initial use of surfactant"

How is this defined? As the first use of surfactant? Or as according to a defined postnatal age?

P6 line 4

Why were septicemia or acidosis at birth not included in the model? Were the data not available in the data set or were assumed that these factors were not important to predict the use of LISA?

P6 line 33

What means "primary endotracheal intubation"? Intubation immediately after birth or during stabilization in the delivery room or during course of RDS?

Results
P7 line 13 Figure 2

I do not find this figure

REVIEWER
Reviewer name: Dr. Amit Mukerji
Institution and Country: McMaster University, Canada
Competing interests: None

REVIEW RETURNED
31-Mar-2022

GENERAL COMMENTS
Major Comments
Intro:
1. When referencing the OPTIMIST-A trial in second paragraph, please specify that this compared LISA with sham treatment (no surfactant)
2. The hypothesis should be more clearly stated. Given the varying designs of randomized trials and heterogeneous results, not sufficient to say hypothesis is results will be different than results from RCTs.

Materials and Methods
3. I think this section can be generally re-organized to flow better.
4. I would clearly defined the exposure (LISA) and control groups (non-LISA) under a subheading "Exposure and Controls". The general non-LISA should be clearly defined as those who received surfactant following intubation, those that received on Non-invasive support without any surfactant, and those that were intubated but no surfactant. In this section, there should be some description of LISA strategies employed across centres. If no particular national-level policy is followed, the ranges of practices should be described – e.g. could say LISA included use of various types of catheters including… and pre-medications could or could not be administered. Also, it should be specified within the control arm (non-LISA) whether both INSURE and standard MV with an intent to keep babies intubated for longer were included.
5. There are essentially two control groups – the “general” non-LISA group, and the non-LISA group that received surfactant. The latter could be labeled different for ease of reading (e.g. non-LISA+S). Could also be placed under separate sub-headings, including in the results.
6. The sensitivity analyses should be with a separate subheading. The infant that was excluded from the analysis due to an unusually long MV duration, should be paced under this sen analysis, and clearly identified to be a post-hoc analysis. And perhaps better fit under results as opposed to methods.
7. Under definition of BPD, replace assisted ventilation with "positive pressure respiratory support"
8. On page 5, it is not clear whether the decision to use PS was an a priori decision, or this was decided post-hoc after reviewing the baseline/demographic characteristics and their distribution?
9. If possible, suggest to provide some references for the choice of variables that were included in the PS model. It would strengthen the basis for the selection of these variables for the reader.
10. Page 6 lines 25-30 – the conduct of a regression model using the same predictor variables as the PS. Is this a model run “within” the PS. I am personally not an expert in PS, but this particular analysis could be described a bit further for the reader. Why do this? Especially if the variables chosen are same as that in the PS model. Is such as analysis better suited as a sensitivity analysis?

Results
11. First paragraphs – the correlations across centres between proportion of LISA and MV duration/proportion – this is not described in the methods as a planned analysis (except brief mention of use of spearman’s correlation for descriptive analysis). If this was indeed a planned analysis, would describe in more detail in the methods. I personally do not think this analysis is necessary as it is probably adding to confusion of the overall message, since the correlation is opposite to the PS results. If choose to keep, need to explore in the discussion why the results are discrepant.
12. The comparisons between LISA and non-LISA and between LISA and non-LISA+S are adequately distinguished, with the latter comparison relegated to the final paragraph. Still, I would suggest to make the distinction of these two sets of comparisons more clear in the text portion of results.

13. As mentioned in comment 10, not clear why the additional analysis (which yields the CI of 0.3-8.0 is needed) and what are these “additional” variables – as it seemed from methods that it would be the same variables as in the PS model. This needs more clarity.

Discussion

14. Suggest to de-emphasize the descriptive results. Start with the PS analysis results.

15. Should separate the first paragraph into two: (1) summary of findings and (2) impact or meaning of findings. Speculations around severity of BPD should not be in the summary of findings but can be expanded upon in the second para. But the second para should be high-level discussion on meaning/impact of these findings.

16. Would include in any comparison with OPTIMIST trial that the interventions studies in that study were different than the present study.

17. If choose to keep the correlations across centres, need to include in the discussion why the results were discrepant and speculate as to reasons why this may be the case.

18. I don’t quite follow the reasoning for the decline in BPD-36 to BPD-40. I think a more reasonable reason might be what was mentioned in 1st paragraph, that the BPD severity may be low overall at 36 weeks, and many of the infants came off support by 40 weeks. I’m not sure how practice variations would account for the drop, since this decline in BPD was seen in both study groups.

19. Agree with the notion that babies in the non-LISA group may have included more sick infants, which may remain unaccounted for despite PS matching. Even though it is mentioned, I would emphasize this by more clearly stating this first in the list of limitations. Currently this concept is interrupted by the mention of BPD-36 vs BPD-40

Tables

20. I think the tables can be re-organized to be easier for the reader. Specific comments below.

a. In Table 1, the means and SDs should be combined into 1 column and presented as means (SD). I also note that in many instances, the SDs are quite high, would it be more appropriate to present these are medians (IQRs). Also the left main columns presents LISA, MV and Surfactant. Why not present as LISA, non-LISA and non-LISA+S to be consistent with the rest of the paper? I frankly think this whole table can just go into supplementary material

b. Table 2 in unnecessarily detailed, esp for main body of manuscript. The last major column (showing data for 25-27 weekers) can be placed in manuscript, and rest can all go into supplementary material. Also, the “characteristics” include both baseline/demographic variables and outcome variables. Eliminate outcome variables from here as they are repeated in Table 3 anyway.

c. I would suggest Table 3 be broken up – one table can be unadjusted (current panel 1) and the other two panels can be in a separate table that shows PS data. Variables used to generate the
PS model should be included in the table footnote.

d. Table 4 can either be simplified or just placed as a supplementary file and mentioned briefly in the results. It is not surprising that variables are similar after PS model was developed using same variables (except BW)

Additional general comments
1. Some grammatical issues throughout document – please review

2. Suggest to limit the manuscript to 25-27 weeks wherever possible. It can be briefly described in the methods that due to the rarity of administration of LISA in babies <25 weeks, these babies were no longer reviewed for any outcomes’ analysis. Data as shown in Table 2 for 22-24 week GA can be shown as an aggregate (combining 22-24 week babies and showing only baseline/demographics) can be added as a supplemental table. In the results, a brief description of the frequency of use of LISA in 22-24 weeks GA can be mentioned.

3. The abstract needs to be clarified a bit further. As written the “Objective” section suggests that non-LISA are only those babies that were intubated. As written, the second section on the correlation across centres is rather unclear (to reader who may only read abstract) – suggest to eliminate this correlation outcome from abstract.

4. Somewhere in manuscript it should be mentioned whether there was any loss to follow-up or whether all outcomes reported are from all patients included in PS models.

Version 1 – Author’s Response

Reviewer: 1 Comments to the Author During the last years less invasive surfactant administration has become more common especially in Europe. Based on some randomised clinical trials and meta-analysis it has been recommended as the preferred mode of surfactant administration in the 2019 European Consensus Guideline. This recommendation was considered weak because some of the underlying studies were open to bias. Furthermore each new method that has been evaluated in trials conducted under standardized and optimized conditions has also to be evaluated under real live conditions. There may be some concerns that become not obvious under standardized conditions but that are important in everyday practice. For example the experience in performing certain skills may influence safety as well as efficiency of a therapeutic method or a method may be beneficial for a certain group of patients but perhaps even harmful in another population. The presented paper by Reigstad and coworkers devotes to the comparison of the outcome after LISA compared to primary endotracheal intubation in infants with a gestational age of 25-27 weeks. They performed an analysis of a defined population-based data set of a national registry (Norwegian Neonatal Network). These data represent the results from a real world scenario. The method of propensity score matching enables some interesting additional results that are worth to be presented and to be discussed. Response: We appreciate the comments on the importance of providing real world scenario data on the
significance of LISA vs. traditional ways of surfactant administration, which is our purpose with the study. However there are some aspects that need revision. Introduction P4, line 42 “It is also a concern that the safe management of LISA requires certain skills, for instance to avoid extended hypoxia during the procedure...” The same could be mentioned for endotracheal intubation. Each method in the hand of an unexperienced person may result in problems. But if a new method has the potential to improve outcome the lack of experience should not be a reason to hinder its introduction. Than it is the task to define each step of the method, to teach it and thereby to increase experience. I propose to discuss the differences in experience with endotracheal intubation on the one hand and LISA on the other hand in different centers as one factor influencing outcome, but not as an advantage or disadvantage of the one or the other method. Response: We agree that intubation also requires certain skills. However, if LISA fails to prevent intubation, the infant will be subjected to two potentially difficult procedures. We would like to stress this challenge and propose the following alteration: “It may also be a concern that a failed trial of LISA and the subsequent need of endotracheal intubation may cause prolonged hypoxia which is known to be associated with ICH (5.) Discussion P8, line 35 It is not correct that only the NINSAPP study (Kribs et al 2015) included infants comparable to those analysed in the presented paper. Also the AMV trial (Göpel et al 2011) included infants with a gestational age of 26 and 27 weeks. Response: We thank the reviewer for pointing out that we overlooked these data. The reference has now been included. P8, line 39 The references (15-24) are obviously not all correct for the statement. Reference 16-19 refer to studies of the INSURE but not of the LISA procedure. Response: These references referred to the references used in an earlier version on the Cochrane «Surfactant therapy via thin catheter in preterm infants with or at risk of respiratory distress syndrome». The references have been updated with the references in the 2021 version. Conclusion The authors state that individual NICU routines may be equally or more important than LISA vs. nonLISA strategy. This may be the case. But it is also possible that under certain defined NICU routines LISA may be beneficial. Theretofor further research should focus on the question which mode of surfactant administration is best under which condition. I think this point should be discussed and perhaps also be mentioned in the conclusion. Response: We appreciate this comment. We have expanded the Conclusion on the reviewer’s suggestion in the updated manuscript and added: It is possible that LISA may be beneficial in certain clinical circumstances, and further research should focus on defining such areas. Some minor points Abstract P3 line 22-23 “and bronchopulmonary dysplasia at postmenstrual age 36 and 40 weeks” As bronchopulmonary dysplasia is defined for a postmenstrual age of 36 weeks it would be better to write: “... need of supplemental oxygen or positive pressure respiratory support” Material and Methods P5 line 28 “initial use of surfactant” How is this defined? As the first use of surfactant? Or as according to a defined postnatal age? Response: For clarity we have changed the wording to: “first dose” P6 line 4 Why were septicemia or acidosis at birth not included in the model? Were the data not available in the data set or were assumed that these factors were not important to predict the use of LISA? Response: The data on septicemia does not clearly define if they were intrauterine (which may be the one of importance) or whether they were postnatal. Furthermore, extensive use of prenatal exposure to antibiotics and therefore lack of positive bacterial cultures make this a difficult parameter. We suspect, however, that severe early onset sepsis also would result in low Apgar scores and CRIB II scores, and in metabolic acidosis. Acidosis is included in the CRIB II score. P6 line 33 What means “primary endotracheal intubation”? Intubation
immediately after birth or during stabilization in the delivery room or during course of RDS?
Response: Changed to “received the first dose of surfactant after endotracheal intubation” Results
P7 line 13 Figure 2 I do not find this figure Response: Included
Review: 2
Comments to the
Author
Major Comments
Intro: 1. When referencing the OPTIMIST-A trial in second paragraph, please specify that this compared LISA with sham treatment (no surfactant)
Response: We appreciate this emphasis – “sham” is now included. 2. The hypothesis should be more clearly stated. Given the varying designs of randomized trials and heterogeneous results, not sufficient to say hypothesis is results will be different than results from RCTs. Response: We respectfully disagree. Our hypothesis is that a result from a real-world experience may differ from results from an ideal RCT. We state the challenges in the two sentences before the sentence expressing the hypothesis and feel that the hypothesis should be “open” to the question of benefits and risks since it is impossible to know if there will be a difference and, if so, in which direction. We therefore suggest that the hypothesis should be left as it is with the exception that “outcomes” replaces “results”.
Materials and Methods 3. I think this section can be generally re-organized to flow better. Response: We have made alterations to make it clearer. 4. I would clearly defined the exposure (LISA) and control groups (non-LISA) under a subheading “Exposure and Controls”. The general non-LISA should be clearly defined as those who received surfactant following intubation, those that received on Non-invasive support without any surfactant, and those that were intubated but no surfactant. In this section, there should be some description of LISA strategies employed across centres. If no particular national-level policy is followed, the ranges of practices should be described – e.g. could say LISA included use of various types of catheters including… and pre-medications could or could not be administered. Also, it should be specified within the control arm (non-LISA) whether both INSURE and standard MV with an intent to keep babies intubated for longer were included.
Response: We appreciate the suggestions and have made a separate subheading “Exposure and Controls”. In this section, we state that there was no national policy on indications or how LISA was performed other than through a thin catheter in the trachea. We now define three groups: the LISA, non-LISA and non-LISA+S groups. This distinction is also adhered to throughout the manuscript, including tables. 5. There are essentially two control groups – the “general” non-LISA group, and the non-LISA group that received surfactant. The latter could be labeled different for ease of reading (e.g. non-LISA+S). Could also be placed under separate sub-headings, including in the results.
Response: We appreciate the suggestions and refer to the response to question 4. We have adhered to these definitions throughout the manuscript. 6. The sensitivity analyses should be with a separate subheading. The infant that was excluded from the analysis due to an unusually long MV duration, should be paced under this sen analysis, and clearly identified to be a post-hoc analysis. And perhaps better fit under results as opposed to methods.
Response: We agree with the reviewer. We have updated the statistics section with a separate paragraph for the sensitivity analyses. However, we find it clearer that each sensitivity analysis is placed next to corresponding main analysis in the results section. 7. Under definition of BPD, replace assisted ventilation with “positive pressure respiratory support” throughout the manuscript. Response: We have changed it accordingly. 8. On page 5, it is not clear whether the decision to use PS was an a priori decision, or this was decided post-hoc after reviewing the baseline/demographic characteristics and their distribution? Response: We have suggested the following introduction to analyses: “First, we assessed if there were correlations between the rates of LISA and rates and mean duration of MV between NICUs. Correlations disclosed differences in background data and treatment policies between NICUs. We therefore used propensity score (PS) analyses to reduce potential bias...” 9. If possible, suggest to
provide some references for the choice of variables that were included in the PS model. It would strengthen the basis for the selection of these variables for the reader. Response: We have included relevant variables that are included in the national register. The same variables may be looked upon as standard variables in this kind of research. We find it superfluous to use a lot of words to explain and document why these variables were included within the limits of words accepted by the journal since we consider them self-evident. We are, however, willing to expand on this if the editor requires it.

Page 6 lines 25-30 – the conduct of a regression model using the same predictor variables as the PS. Is this a model run “within” the PS. I am personally not an expert in PS, but this particular analysis could be described a bit further for the reader. Why do this? Especially if the variables chosen are same as that in the PS model. Is such an analysis better suited as a sensitivity analysis? Response: We do this mainly for three reasons: Firstly, while the PS matching will generally balance confounders and thus eliminate (or at least greatly reduce) bias in the treatment effect estimates, an unadjusted analysis will still only give treatment effect estimates at the population level. If we additionally adjust for the variables using regression, we effect estimates at individual level. Doing this should also remove any residual covariate imbalance. And adjusting for predictors of the outcomes will also theoretically increase the precision (and statistical power) in the estimates. (For the same reason one should adjust for strong predictors of the outcome even in randomised studies.) However, in this study, the estimates from the regression models did not differ very much from the estimates in the unadjusted models, so we give more emphasis to the simpler analyses. We have rephrased and expanded the relevant section in the methods section to more clearly (but concisely) explain this. It now reads: To estimate the expected difference in the number of days on MV at the individual (not only population) level, remove the effect of any residual baseline imbalance and potentially increase the precision of the estimates, we additionally used a linear mixed-effects model with the same explanatory variables as in the PS model.

Results 11. First paragraphs – the correlations across centres between proportion of LISA and MV duration/proportion – this is not described in the methods as a planned analysis (except brief mention of use of spearman’s correlation for descriptive analysis). If this was indeed a planned analysis, would describe in more detail in the methods. I personally do not think this analysis is necessary as it is probably adding to confusion of the overall message, since the correlation is opposite to the PS results. If choose to keep, need to explore in the discussion why the results are discrepant. Response: The correlation analyses was an a priori decision to examine if there was a correlation between the frequency of LISA and frequency and duration of mechanical ventilation. We have expanded on this issue early in the Statistics section and in the Result and Discussion sections. We find it interesting that there were marked differences in the use of LISA, and in fact, a positive correlation between the frequency in the use of LISA and frequency and duration of mechanical ventilation between institutions. We suggest that the opposite finding in correlation analyses and the PS model effectively illustrates the need for models like the PS model when comparing institutions.

12. The comparisons between LISA and non-LISA and between LISA and non-LISA+S are adequately distinguished, with the latter comparison relegated to the final paragraph. Still, I would suggest to make the distinction of these two sets of comparisons more clear in the text portion of results. Response: We agree and refer to our response to #5. This also includes the tables. 13. As mentioned in comment 10, not clear why the additional analysis (which yields the CI of 0.3-8.0 is needed) and what are these “additional” variables – as it seemed from methods that it would be the same variables as in the PS model. This needs more clarity. Response: We thank the reviewer for this comment. Please, see our response to # 10. Discussion 14. Suggest to de-emphasize the
descriptive results. Start with the PS analysis results. Response: We find it necessary to describe a comprehensive “take-home message” since the results may be difficult to grasp first-hand from the Results section. We suggest to start with the correlation analyses since we now have described why these analyses were performed, and in this first paragraph of the Discussion it illustrates the need for doing analyses like the PS in multicentre cohort studies. We have, however, shortened and simplified the paragraph substantially. 15. Should separate the first paragraph into two: (1) summary of findings and (2) impact or meaning of findings. Speculations around severity of BPD should not be in the summary of findings but can be expanded upon in the second para. But the second para should be high-level discussion on meaning/impact of these findings. Response: We have divided the paragraph and discussed accordingly. 16. Would include in any comparison with OPTIMIST trial that the interventions studies in that study were different than the present study. Response: We appreciate this distinction and have stated that the design of the OPTIMIST trial is different. 17. If choose to keep the correlations across centres, need to include in the discussion why the results were discrepant and speculate as to reasons why this may be the case. Response: We have expanded on this issue in the second paragraph of the Discussion. 18. I don’t quite follow the reasoning for the decline in BPD-36 to BPD-40. I think a more reasonable reason might be what was mentioned in 1st paragraph, that the BPD severity may be low overall at 36 weeks, and many of the infants came off support by 40 weeks. I’m not sure how practice variations would account for the drop, since this decline in BPD was seen in both study groups. Response: We respectfully disagree since BPD at 36 weeks may include significant differences according to, for instance, acceptance of hypercapnia and differences in oxygen saturation. We wish to argue that the marked and similar decrease in need of oxygen supplementation or positive pressure respiratory assistance from PMA 36 to 40 weeks underscores that the severity of BPD was probably the same for the two groups. Since severity of BPD is a prime argument for proposing LISA, this is an overriding important message of this study. 19. Agree with the notion that babies in the non-LISA group may have included more sick infants, which may remain unaccounted for despite PS matching. Even though it is mentioned, I would emphasize this by more clearly stating this first in the list of limitations. Currently this concept is interrupted by the mention of BPD-36 vs BPD-40 Response: We agree that this is a potentially very important confounder and have suggested to move this discussion to the second paragraph in the Discussion (discussion of the significance of findings) and only briefly mention this weakness due to unmeasured potential confounders in the strengths and weaknesses paragraph. Tables 20. I think the tables can be re-organized to be easier for the reader. Specific comments below. a. In Table 1, the means and SDs should be combined into 1 column and presented as means (SD). I also note that in many instances, the SDs are quite high, would it be more appropriate to present these as medians (IQRs). Also the left main columns presents LISA, MV and Surfactant. Why not present as LISA, non-LISA and non-LISA+S to be consistent with the rest of the paper? I frankly think this whole table can just go into supplementary material Response: We have made some alterations according to the suggestions, like putting SD in parenthesis without separate columns. However, the importance of the table is to visualize the national cohort and how routines differ between NICUs. b. Table 2 in unnecessarily detailed, esp for main body of manuscript. The last major column (showing data for 25-27 weekers) can be placed in manuscript, and rest can all go into supplementary material. Also, the “characteristics” include both baseline/demographic variables and outcome variables. Eliminate outcome variables from here as they are repeated in Table 3 anyway. Response: We agree that the table is extensive, but tables tell the detailed story better than any text. We therefore suggest to keep the table, but are willing to delete the columns for the
gestational ages below 25 weeks, and only include the individual columns for GA 25, 26 and 27 and the combined 25-27 and the statistics column. We ask the editor to decide. c. I would suggest Table 3 be broken up – one table can be unadjusted (current panel 1) and the other two panels can be in a separate table that shows PS data. Variables used to generate the PS model should be included in the table footnote. Response: We respectfully disagree. To keep all the tables together flows well with the text and avoids repetition of descriptions of variables. In our view a table, no matter how large, is easier to read than text. But again, we are willing to change if the editor recommends a change. d. Table 4 can either be simplified or just placed as a supplementary file and mentioned briefly in the results. It is not surprising that variables are similar after PS model was developed using same variables (except BW) Response: Table 4 is included to show that the PS model was adequate, i.e., that it successfully balanced the baseline variables. We have tried to follow a proposed reporting guideline for propensity score analysis papers (https://doi.org/10.1093/jnci/djw323), which requires that this information is given (checklist items 28 and 29). Additional general comments 1. Some grammatical issues throughout document – please review Response: We have made alterations throughout 2. Suggest to limit the manuscript to 25-27 weeks wherever possible. It can be briefly described in the methods that due to the rarity of administration of LISA in babies

VERSION 2 – REVIEW

| REVIEWER | Reviewer name: Dr. Peter Flom Institution and Country: Peter Flom Consulting, United States Competing interests: None |
|---|---|
| REVIEW RETURNED | 01-Jun-2022 |
| GENERAL COMMENTS | I confine my remarks to statistical aspects of this paper. The general approach is fine, but I have some issues to resolve before I can recommend publication. One general question - and I am not entirely sure of the right approach - is whether the authors should do the traditional sort of hypothesis test (null = no difference) or a test of equivalence (TOE) (null = a difference greater than XXX). The answer depends on the goal of the research. The reason I raise this at all is the name of LISA (less invasive). If it is known that LISA is preferable (because less invasive methods are generally preferable) and you want to see if it is "just as good" as non-LISA, then a TOE should be considered. But if your goals are more traditional (which one works better?) then the methods used are correct. |
| | p. 4 line 27 - insert "significant" between "no" and "association" (unless the result really was exactly the same). (The same issue occurs at some other places e.g. p. 5 line 32). |
| | p. 7 line 20 Using both GA (as a continuous measure) and SGA vs. non-SGA seems like it could cause problems such as collinearity. Why include the dichotomous measure, especially as you (quite properly) used splines to look at nonlinearity. |
| | line 42 Welch's t-test assumes the data are normally distributed. Table 2 makes it clear that this isn't close to correct. I suggest a more robust measure |
| | Table 2 - the "days" variables should be summarized with median and either interquartile range or median absolute deviation. The SDs are often larger than the means (or close to the same size) and these variables can't be negative. |
Version 2 – Author’s Response

Associate Editor June 17th, 2022

Comments to the Author: What is already known

1st statement add "and subsequent surfactant administration' after "intubation" Done.

Add a The Patient and Public Involvement statement to the text (see Instructions to authors) Done.

Reviewer: 1 Dr. Peter Flom, Peter Flom Consulting

Comments to the Author I confine my remarks to statistical aspects of this paper. The general approach is fine, but I have some issues to resolve before I can recommend publication.

One general question - and I am not entirely sure of the right approach - is whether the authors should do the traditional sort of hypothesis test (null = no difference) or a test of equivalence (TOE) (null = a difference greater than XXX). The answer depends on the goal of the research. The reason I raise this at all is the name of LISA (less invasive). If it is known that LISA is preferable (because less invasive methods are generally preferable) and you want to see if it is "just as good" as non-LISA, then a TOE should be considered. But if your goals are more traditional (which one works better?) then the methods used are correct. Yes, our goals are more 'traditional'.

For neonates with low gestational age it is not established whether LISA or intubation and surfactant treatment is preferable. p. 4 line 27 - insert "significant" between "no" and "association" (unless the result really was exactly the same). (The same issue occurs at some other places e.g. p 5 line 32). We have added the word 'significant' where appropriate. p. 7 line 20 Using both GA (as a continuous measure) and SGA vs. non-SGA seems like it could cause problems such as collinearity. Why include the dichotomous measure, especially as you (quite properly) used splines to look at nonlinearity. We are very sorry, but the text referring to SGA was accidentally included in the method section, left from an earlier draft. Our PS model did not in fact include the SGA variable. (An earlier model (with very similar results) did, but we excluded it after including the more detailed CRIB II score, which already includes SGA as one component.) In any case, SGA was similar both before after matching by PS. We have updated the text. While collinearity should not really be a problem in propensity score models, we did check for collinearity in all models fitted, and there were no indications of collinearity problems (e.g., all generalised VIF values were < 2). line 42 Welch's t-test assumes the data are normally distributed. Table 2 makes it clear that this isn't close to correct. I suggest a more robust measure For these data, the number of observations is so large that Welch's t-test works very well (the distribution of the means are well approximated by a normal distribution, due to the central limit theorem), i.e., the test is robust. It is powerful enough to detect (even small) differences between groups – as witnessed by the many small P-values – while still preserving the type I error rate. (We tested the latter by running t-tests on the data with simulated permuted/random group allocation.) For the main outcomes, we also calculated bootstrap-based confidence intervals, but they were almost identical to the confidence intervals from the t-test, so to keep presentation simple for the reader and to keep the number of words down, we only report the results from t-test. For example, for the difference in number of days on MV in the unadjusted analysis, the 95%
confidence interval (CI) from the t-test was 0.8–5.2, while the corresponding percentile (and BCa) bootstrap CI was 0.9–5.2. Table 2 - the "days" variables should be summarized with median and either interquartile range or median absolute deviation. The SDs are often larger than the means (or close to the same size) and these variables can't be negative. Due to the quite discrete nature of the ‘days’ variables (e.g., 24% of the patients have 0 days, 7% have 1 day), the value of the median isn’t very ‘stable’ (i.e., adding or removing a patient can easily increase/decrease the median by one whole day), so we prefer the mean as our main outcome measure. However, as the reviewer correctly points out, the SDs are large compared to the means, and more information on the distribution of ‘days’ may be of interest. We have therefore now added the median and the lower and upper quartiles as an additional row for each ‘day’ variable in Table 2. For the variables with significant differences, the overall pattern is the same, but, as expected, due to the right-skewed distributions, the medians are smaller than the means. (A note to the BMJ editors - putting line numbers and so on on the pages with figures makes them hard to read. This is especially true for figure 2, where the line numbers overlap some key numbers, and the page number and other text is also distracting. It would be good to fix this. But this has nothing to do with the current paper in particular). Overall, a good job of a complex analysis. Thank you! Yours sincerely Hallvard Reigsta

| REVIEWER | Reviewer name: Dr. Peter Flom  
Institution and Country: Peter Flom Consulting, United States  
Competing interests: None |
|----------|--------------------------------------------------------------------------------|
| REVIEW RETURNED | 17-Jun-2022 |
| GENERAL COMMENTS | The authors have addressed my concerns and I now recommend publication  
Peter Flom |