Peer Review File

Article information: https://dx.doi.org/10.21037/jhmhp-23-71

Reviewer A

Comment 1: This manuscript reports on the effects on LOS of a change in how patients were assigned to care teams in the NICU so that the lower acuity patients and higher acuity patients were managed by separate care teams. They report that this is associated with a clinically meaningful reduction in LOS. While I have concerns that are noted below, if the findings hold up this is a clinically important result. I will note that my prior is that the results will hold up as I know of NICUs that don't do this sorting where the lower acuity patients do not receive all of the clinical attention that is needed, and that at least some of the clinicians in these units believe that this does result in unnecessary increases in LOS.

Reply 1: Thank you for your comment and your response on contributing factors to LOS.

Changes in the text: n/a

Comment 2: The details about the study were very sparsely explained and there are places where details are lacking, so I am going to first summarize what I think the authors did. All of my comments are based on what is outlined below.
The intervention appears to be that a scoring system was used to assign infants to one of 3 care teams, two of which cared for higher acuity infants, and the other cared for lower acuity infants. Higher acuity infants would normally transition to the lower acuity team when their acuity reduced (e.g., VPT infants who had advanced to grower-feeder status).

Reply 2: Yes, that is correct. Higher acuity infants would transition to lower acuity team when acuity was reduced as example you listed there with VPT infants advanced to feeder grower status.

Changes in the text: We have expanded the description of the intervention in the NICU team structure of the Methods Section, including clarification regarding higher acuity infants transitioning to the lower acuity team when approaching discharge. We have included the example of VPT infants advancing to grower-feeder status on Page 9 Lines 152-154.

Comment 3: The authors used a regression analysis to assess the effects of this intervention on risk-adjusted LOS.
It is not clear how care was organized prior to the intervention. It isn't stated, but I am inferring the prior to the intervention that patients would be assigned to a care team without regard to acuity, and that team wouldn't change before discharge. Regardless of what the change was, it needs to be described.

Reply 3: Thank you for requesting further clarification. That is correct regarding the pre-intervention team structure.

Changes in the text: Added line with clarification point requested above on page 8 lines 134-137.

Comment 4: I will note that the post-intervention care process mocks what many NICUs have been doing for a long time; sorting the NICU into a NICU-level space and an intermediate NICU space, with separate care teams for each. I personally know of at least one NICU that had a separate intermediate unit dating back to the advent of NICUs in the mid 1960s. There is variation in how this structure is managed; in some hospitals staff always or primarily are assigned to one of the levels, while the staff may rotate across teams over time, but for any given shift they are assigned to a specific team/level of care.
Assuming that my understanding of what the change was is correct, the findings are not surprising as if the team is caring for infants at different levels of acuity, the team attention will be pulled away from the lower acuity infants.

Reply 4: Thank you for your comments and reference to other hospital structures and staffing models. Yes, that is correct with the concern for attention being pulled away from lower acuity infants.

Changes in the text: Added lines describing this structure with reference to other hospital layouts with dedicated step-down units on page 8 lines 141-143.

Comment 5: This study has value as comparing the two models of care on LOS is something that has not been well studied. Further, given that it is very unlikely that anyone would ever conduct a randomized trial comparing the two care models, the analytical approach used is reasonable. Ideally it would be done over several different hospitals to limit the threat to validity of other unobserved that overlapped the change in model of care. But the single unit study reported has value.

Reply 5: Thank you for your comment. We listed single center study in the limitations section of discussion.

Changes in the text: Expanded the conclusions on the need for future research to test the bed aggregation intervention across multiple sites to limit threat to validity page 14 lines 271-272.

That said, I have significant concerns that need to be addressed.

Comment 6: Major: The details provided are grossly inadequate. It is not possible to fully assess what was done. With the caveat that the sparse details about the methods may have caused me to not understand what the authors actually did, it appears that there are serious issues with the methods that are noted below.

Reply 6: Thank you for your comments. The inclusion criteria are listed in the definitions of measures section. Will address remaining comments below.

Changes in text: added evaluating LOS to methods section page 7 line 124.

Comment 7: Major. I think that the regression model is inadequate/not properly specified. Several changes are needed, in no specific order: While I don't want to make a conjecture about how it is affecting the results, the errors in the model are clearly going to introduce a lot of prediction error. This could result in either an under- or over-prediction of the treatment effect.
Comment 8: It isn't stated, but I am assuming that inclusion of GA in the model was by week. If not, this needs to adjusted or justified (I acknowledge that some aggregation may be needed for the lower GA). Assuming that this is correct, why was BW also included in the model? BW and GA are highly correlated. Plus, once GA is controlled for what matters is if the infant is SGA, and non-SGA differences in BW aren't that significant. A better model would be to replace BW with an indicator for SGA. The authors might consider testing to allow the SGA effect to vary with GA (e.g., the effect of being SGA is different for the lowest GA).

Reply 7&8: We have included revised the regression models to address these concerns, including the following: (1) in the overall model, GA is included as a continuous variable, as bivariate plots of GA and LOS show a relatively linear relationship; additionally, a binary variable for <33 weeks GA is included to indicate infants that would likely be seen by the higher acuity care team; (2) dropped birth weight from the models; and added small for gestational age at birth. We have stratified our regression models by GA (<32 weeks, 32-33 weeks, 34-36 weeks, ^{3}37 weeks), which addresses heterogeneity in the occurrence of SGA by GA.

Changes in text: See methods section lines page 9 164-167, page 11 198-201. Added line on prediction error to discussion section page 13, lines 254-256.

Comment 9: There is just one indicator variable for "acuity" which is essentially a dummy variable for if the infant was in the high acuity group. This is totally inadequate for a model that is trying to predict LOS. The marginal effect on LOS of the conditions included in the acuity indicator are very heterogenous. Especially given the relatively small sample, this is inducing a lot of prediction error. While the sample size precludes careful model development of the effects of each of the variables in the acuity score, the authors should use some clinical judgement to make logical groupings and then possibly use a preliminary round to revise these groupings.

Reply 9: We have added several variables to the regression models to better capture patient acuity that would likely require care by the higher acuity care team, either at the time of birth or during the NICU stay, including the following measures: (1) <33 weeks gestational age as those infant typically require respiratory support at birth ranging from CPAP to intubation; (2) other high acuity diagnoses at infant birth; (3) the occurrence of one or more neonatal complications during the NICU stay, including necrotizing enterocolitis, retinopathy of prematurity, or intraventricular hemorrhage,; and (4) patent ductus arteriosus which can cause hemodynamic instability. Additionally, to address the possible heterogeneity across GA, we have stratified the analysis by GA (<32 weeks, 32-33 weeks, 34-36 weeks, ^{3}37 weeks). We found that substantial reductions in LOS occurred in infants ^{3}37 weeks GA.

Changes in Text: Page 9 lines 168-173. This finding is reported in the Results and described in the Discussion.

Comment 10: I didn't carefully review all of the codes that were included in the acuity indicator. But I noted some that probably shouldn't be included. For example, ASD and VSD probably shouldn't be included. There are a large number of ASDs/VSDs that are diagnosed, but are small and not clinically meaningful. Further, essentially all clinically meaningful ASDs/VSDs are in conjunction with another, clinically meaningful cardiac anomaly. Those other anomalies are what should be included in the indicator. The added effect of adding ASD/VSD to the acuity indicator on for the defects that are not clinically meaningful will induce bias. Either in the text, the appendix, or both, information needs to be added about how the ICD codes were selected to be included in the acuity indicator.

Reply 10: Thank you for your comment. We included ASD and VSD as infants with this condition can have respiratory challenges and feeding difficulties leading to ICU admissions and longer LOS. We noted in the definition of measures section acuity was determined by neonatologists.

Changes in the text: We have revised our measures of acuity and now include four variables. The high acuity diagnoses at birth were determined by neonatologists and include diagnoses that would likely require the infant to be cared for by the high acuity care team at birth (Page 9 lines 167-173). Added a statement that a group of neonatologists determined the list of diagnosis for higher acuity infants to page 10 lines 174-178. As described in Reply 9, other measures of acuity include <33 weeks gestational age; (3) the occurrence of one or more neonatal complications during the NICU stay, including necrotizing enterocolitis, retinopathy of prematurity, or intraventricular hemorrhage; and (4) patent ductus arteriosus.

Comment 11: The authors present one unified effect estimate, but it is likely that the effect varies quite significantly across the range of possible effects. Their exclusion of infants with very short stays does help with this as a very significant share of the infants who are admitted to the NICU have an LOS<6 days, especially among term and near-term infants. Even so, it is impossible for an infant who previously had a 7-day LOS to experience the observed mean of a 5 day reduction in LOS. At a minimum the authors should conduct some simple tests to examine if the intervention effects vary by GA.

Reply 11: We have made multiple changes to the regression models, including stratifying the models by GA (<32 weeks, 32-33 weeks, 34-36 weeks, ^{3}37 weeks).

Changes in Text: As noted above in 7-10.

Comment 12: Minor, possibly more significant. It isn't clear how acuity was defined. The appendix lists that codes that were considered, but this is not the universe of possible codes. There are a fair number of similar severity codes that were not included. I am inferring from this that the list provided was limited to those that were in the data for the study period. This should be made clear. I understand that clinical judgement was the actual criteria for who were the higher acuity patients.

Reply 12: yes, the codes were limited to those in these data for the study period and clinical judgement was actual criteria

Changes to text: added clinical judgement to page 10 line 175. Added limitation of codes available during study period to page 10 lines 177-178.

Comment 13: Minor: Table 1. GA should be reported. As noted above, GA, not BW is what drives clinical management.

Reply 13: We have made multiple changes to the regression models, including stratifying the models by GA (<32 weeks, 32-33 weeks, 34-36 weeks, ^{3}37 weeks).

Changes in text: table 1 reports GA as a continuous variable (median) and GA categories. Additionally, Table 1 reports the proportion that are small for gestational age at birth.

Comment 14: Minor. The only serious morbidity reported is BPD. Assuming that the GA distribution of the reported BW groups is similar to population norms, the sample will have well over 200 VPT infants. From large population-based data, the rate of mortality or major morbidity (BPD, ROP, IVH, or NEC) for VPT infants is about 50%, with just under 20% dying. This would imply a fairly large number of infants in the sample with one or more of these morbidities. While the numbers may be smaller for IVH, NEC, and ROP, at least the combined number should be reported. These complications have an effect on LOS, and these data are needed to put the results in context.

Reply 14: We have replaced BPD with a variable to indicate one or more neonatal complications (BPD, necrotizing enterocolitis, retinopathy of prematurity, intraventricular hemorrhage) versus no neonatal complications. Due to low prevalence of NEC, ROP and IVH, we were not able to includes these complications in the model as separate variables.

Changes in text: As noted in methods section page 9 lines 171-172.

Reviewer B

Comment 1: I will acknowledge my limited knowledge of statistical analysis upfront in advance of my comments, but note that this will be true also of most readers, so perhaps the article just requires a better explanation for some of the items of my concern.

Reply 1: thank you for your comment. Will address items below as listed.

Changes in text: n/a

Comment 2: Confusion #1 - the authors state in lines 190-191 that "there were no differences in infant or maternal demographic characteristics, family visitation, or presence of BPD", but then uses these same factors in line 195 to show that a non-significant difference in LOS prior to adjustment then becomes significant when adjusted for these factors.

Reply 2: We have clarified in the text that there were no statistically significant differences between aggregation periods in these characteristics in the bivariate analyses, demonstrating that infants in the two aggregation periods were similar. We have included these variables in the regression models, as they are theoretically associated with LOS.

Changes in text: Results page 11 lines 206-209, 215-217.

Comment 3: Confusion #2 - an average LOS of 5 weeks in a population of infants whose median GA is 34 weeks is unusually high and suggests some opportunity for improvement certainly existed. Using a different staffing model to accomplish a 5 days reduction in the LOS is striking, and again suggests that in the pre-aggregation period that many babies were staying longer than necessary. The authors offer no insight as to the ways in which this improvement was achieved - were babies advanced more rapidly to all oral feedings? Were they weaned from respiratory support more quickly? Were families prepared better for their care at discharge? Understanding the mechanism by which this team approach accomplished such a major change in LOS is important, but the authors offer no insight.

Reply 3: Thank you for your comment. We added information on the decreased length of stay contributing factors.

Changes in the text: Added potential contributing factors to decreased LOS as mentioned above in discussion page 12 lines 227-230.

Comment 4: The authors also offer a few head fakes in their introduction and discussion. In the intro they discuss NICU design changes to single family rooms in the second paragraph and noise pollution in the third paragraph, yet their strategy of assigning care teams by acuity level has no obvious connection to either of these observations. NiCU layouts are again mentioned in the conclusion without any clues as to how the authors' think their research into care team assignment could inform NICU design. In old open-bay NICUs, one room was often designated as a "step-down" area, achieving a similar model of a dedicated nursing staff to the convalescing babies; in the authors' single family room NICU they say that they didn't study "the physical movement of patients throughout the NICU hospitalization", yet it seems the only way to cluster nursing assignments would be to have the low-acuity patients physically proximal to one another, which would certainly require movement of some babies during the hospitalization.

Reply 4: Thank you for your comments. Yes, nursing assignments were based on acuity level as well, see additional information added to text as noted below.

Changes in text: Added comment on historical open bay units with step down facility to intro section page 6 lines 96-99. Added comment on clustering infants by acuity level and restful states to intro page 7 lines 106-108. Added summary of acuity location and assignments to conclusions page 13-14 lines 263-268. Added line on clustering nursing assignments and proximity to page 13 lines 258-259.

Comment 5: As the authors note, the study used historical controls which is a very possible confounding factor. One can imagine that the lessons learned in how to get babies discharged more quickly might be retained even if the nursing teams were no longer assigned to keep high and low acuity infants in separate teams.
Any study that can show a significant reduction in LOS is of interest so this one is certainly worthy of consideration, but without identifying the ways in which dedicated teams achieved the reduced LOS the reader is left to guess. The authors do not help with those guesses by offering their own explanation of how the days to discharge were reduced but rather point the readers to NICU layout and noise levels - neither of which is likely to have been the reason these gains were achieved since the intervention did nothing to address either factor.

Reply 5: Thank you for your comment. Yes, historical controls might have been a confounding factor. Added potential explanations for decreased LOS with more time available to dedicated to discharge tasks and education.

Changes to text: Infants are balanced between the aggregation periods, with no significant differences in any infant and maternal characteristics. This suggests that differences in LOS are not due to differences in these characteristics. Added historical controls as possible confounding factor to limitations section see page 13 line 256. Added potential explanations for decreased LOS to conclusions section page 13-14 lines 263-271.