PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Nutritional practices and growth of preterm infants in two neonatal units in the UK and Malaysia: a prospective exploratory study
AUTHORS	Abdul Hamid, Haslina Szatkowski, Lisa Budge, Helen Cheah, Fook-Choe Ojha, Shalini

VERSION 1 – REVIEW

REVIEWER	Reviewer name: Dr. Sahana Devadas Institution and Country: not applicable Competing interests: none
REVIEW RETURNED	28-May-2021

GENERAL COMMENTS	The sample size could have been more.

REVIEWER	Reviewer name: Dr. Peter Flom Institution and Country: Peter Flom Consulting, United States Competing interests: none
REVIEW RETURNED	11-May-2021

GENERAL COMMENTS	I confine my remarks to statistical aspects of this paper. The general approach is appropriate, but I have a number of issues to resolve before I can recommend publication. page 6, line 47-55 So, it seems like these hospitals are not really comparable, being at different "levels." Shouldn't similar level hospitals be chosen? page 7, lines 11-15 The number of subjects should be based on power analysis. Was one done? If not, then state that this was simply a convenient number to collect. General: For "age" (as in weight for age) was gestational age used, or was it time since birth? page 8, line 48 and other places. What tests were used? Please state here how you decided, and, when each p value is mentioned (in text or table) please state the test. In some cases, t-tests are clearly not right as the data are not close to normal. line 52-53 This is known as bivariate screening and it is not a good method. It can be shown that all the output from the multivariable equation is wrong (standard errors are too small, p values are too low, parameter estimates are biased away from 0; see Frank Harrell's book Regression Modeling Strategies for details). It's better to use substantive knowledge, but if an automatic method must be used, LASSO isn't bad. line 58-60 Replace "correlation" with "collinearity" -- they are not the same.

	p. 10 line 35 and other places -- insert "significant" between "no" and "difference". p. 11 line 37-40 Don't rely too much on significance. Look at effect size. Table 1 - What are all the numbers? E.g. for gestational age, there are two sets of numbers in parentheses. One of them (but which one?) is, I assume, IQR. But what is the other one? Table 3 - what is MD? I'm guessing it is mean difference, but please state this. Peter Flom

REVIEWER	Reviewer name: Dr. Linda Adair Institution and Country: not applicable Competing interests: none
REVIEW RETURNED	08-Jun-2021

GENERAL COMMENTS	This is an interesting comparative analysis of preterm infants born in two very different settings. While there is much useful information presented, there is insufficient explanation of many key variables, and lack of clear perspective for interpretation of results. Additional information needs include: 1. What are the hospitals' criteria for discharge (which may have influenced length of stay)? 2. In estimating the nutritional content of breast milk, were gestational age and time postpartum taken into consideration? 3. It is stated that "Postnatal growth failure was defined as a decrease in WAZ between birth and discharge of ≥ 1.28 as used in previous studies". Both of the cited studies used a weight Z-score of <-1.28 at discharge as a criterion for growth failure, while Lin et al also use a change in Z-score of -1.28. The latter represents quite significant weight loss, while the former may represent a persistently small infant who tracked below the 10th percentile for weight. It would be helpful for readers to better understand the degree of weight loss (in grams or % birth weight) represented by this definition of growth failure. 4. Information about the method of assessment and reliability of gestational age should be presented. 5. Rare clinical conditions are noted as those that occurred in fewer than 10 babies per site, but this is 20% of the sample in each setting and thus not really rare. 6. In table 2 values are show for the proportion of energy from breast milk and formula but does not indicate the time period this represents 7. Details are needed on how cumulative energy and nutrient deficits or excesses were calculated. ESPGHAN reference values are presented in ranges: what values were actually used? Are these appropriate for the full range of gestational ages. 8. Table 4 should present sample sizes for each analysis 9. Is there NO feeding of infants at the breast? Was all breast milk pumped and fed to infants? If not, how was intake estimated? 10. In table 3, changes in weight Z-score should be adjusted for age duration of stay. Analysis and Interpretation 1. In the statistical analysis section, it is stated that univariable analyses were used to explore factors that predicted changes in WAZ". Do you mean bivariable (associations of each risk factor with the outcome, one at a time)? 2. The potential for selection bias related to missing data from earlier discharge should be more thoroughly discussed, since larger, healthier infants were likely discharged earlier and analyses such as those on nutritional intakes in weeks 1-4 will exclude the early

	discharges. Alternately, the analysis in table 4 could be restricted to intake in the first 2 weeks. 3. Differences in nutritional intakes will result from the balance of feeding practices. The differences could be explained with analysis of the feeding data. Based on the results, it is clear that data on types and amounts of feeds are available. Thus, instead of speculating about the distinctions in practice, an important contribution would be to analyze how the different feeding strategies relate to nutritional intakes. 4. Postpartum weight loss and regain is expected in all newborns. Time to regain birthweight is shown in table 3 but not discussed in the text. Moreover, other studies show that birthweight regain is related to mode of delivery: another variable that is shown in table 1 not discussed in the paper. 5. The objective of the study was to compare nutritional practices across neonatal units in very different settings, and the paper presents a good description of the differences, but offers little insight into what accounts for the differences. The second aim is to investigate the association of nutritional practices to postnatal growth. This requires a clear conceptual model. While the focus is on associations, not causal inference, there should nonetheless be some discussion of the direction of associations. A perspective that is lacking in the paper is that feeding practices are both a response to growth and other health conditions AND a predictor of growth. Change in feeding strategies may occur if an infant is not gaining weight. The discussion should acknowledge these important relationships and their implications for interpretation of the results.

VERSION 1 – AUTHOR RESPONSE

Comments	Responses	Changes in main document
Clinical reviewer		
Title: "adding nutritional practices is more appropriate"; 'title looks misleading. adding that its a two center will be apt'	Title has been changed to: **Nutritional practices and growth of preterm infants in two neonatal units in the UK and Malaysia: a prospective observational study**	Change on title (page 1)
What this study adds. Bullet item one "already known";	This point has been added to “what is already known”	Changes on page 4

bullet item three "this is a recommendation and the study does not seem to have subgroup analysis of the lower GA "	The sentence has been amended to those at-risk of severe co-morbidities.	Changes on page 4
bullet item 4: "need to address if its in a manner benefiting or harming."	The sentence has been amended to “postnatal growth failure”	Changes on page 4
Methods, first paragraph: "why were the centers chosen?	1. Malaysia, in general as compared to other south east Asian countries, has similar government funded healthcare as in the UK and a low neonatal mortality rate, indicating availability of specialist neonatal care. 2. Both neonatal units in Malaysia and the UK are based in teaching hospitals, were chosen due to the convenience of data collection with the availability of collaborators and research assistants and prompt ethics approval at the Malaysian site 3. Though different levels of care in	Changes on page 5

	general due to the absence of inhouse surgical support in the UK unit, both units have similar range of preterm infants admission per year of 360-600 infants (monthly 30-50 infants), number of beds (25-26) and offer care at 3 levels: intensive care, semi-intensive and convalescent.	
How can the practices generalizable to the respective countries?	The 2 units studied are broadly representative of neonatal units in their respective countries. Feeding practice such as the use of breastmilk and BMF are very common in other MAL units as well. As for the UK, the rate of breastmilk feeding at discharge is comparable to the audit report by NNAP – indicating the generalised use of breastmilk in the unit. Many units in both countries adopt their own feeding protocols, but ESPGHAN is mostly cited as the basis of these feeding protocols.	-

how are the 2 centres similar in providing care?	Except for not providing inhouse surgical support in the UK unit, both units provide similar types of care which comprises of care for the stable to intensive care infants.	Added this information on page 5
authors should have included a demographic statistic of respective NICUs for comparison. "	Thank you for your suggestions. I have added this information on page 5	Added this information on page 5
Methods, second paragraph: "why was the sample restricted to 50? calculation of sample size to show significance?"	The sample size was determined based on the usual admission numbers and length of stay at the neonatal units of both countries, with the aim of ensuring that daily data collection from birth to discharge was feasible for one person collecting data. For both neonatal units, usual monthly admissions of preterm infants range from 30 to 50 infants. Therefore, collection of data from 50 infants from each unit was deemed to be feasible and achievable.	-As given on page 6- the sample size was restricted due to the limit on study duration – the sample size of 50 “was determined based on the usual monthly admissions and length of stay at the respective units”
Results, first paragraph: "since the birth weight is also an independent predictor for EUGR and neonatal morbidities, it is	We compare mothers’ age, parity and multiple birth in this study. We did not have data on other relevant variables and	-

important to compare the prenatal factors responsible for SGA."	have mentioned this as limitation in this study on page 13 and page 14.	
Results,enteral feeding paragraph : "It would be good to describe if there were any standardized feeding protocols being followed in the units and their basis.	This has been added, thank you.	Changes has been made on page 7
the maximum fluid intake in both the cohorts seem less than the recommendation (160-180ml/kg) even through discharge. was the peak weight loss in infants looked at?"	Yes. Weight loss in all infants peaked at week 1 of admission	- This information has been added on Table 3
Discussion, second paragraph: "the enteral feeding initial=n and progression are no different. is it possibly due to the local PN protocols?"	The number of infants started on PN are different but those started on enteral feeds are similar possibly because in the UK unit, infants were started on enteral feeds with a view to advance to full milk feeds quickly and hence not given PN, in keeping with local protocols. In the Malaysian unit, infants were sicker, and more SGA. Although they were	

	given some enteral feeds, they were also started on PN with the anticipation of a greater risk of feeding intolerance and a slower advancement of milk feeds with a view to boost nutrition with PN in the meantime.	
Discussion, third paragraph, typo: "This may be as due"	Thank you. The amendment has been made.	Changes made on page 12.
Conclusion: "the conclusion need to concur with the primary and secondary objective. may be more specific."	Thank you. More specific conclusion has been made.	Changes made on page 14.
Reviewer 1		
page 6, line 47-55 So, it seems like these hospitals are not really comparable, being at different "levels." Shouldn't similar level hospitals be chosen?	Although the units are categorised differently due to the classification in the categories of neonatal care between UK and Malaysia, the level of care is similar aside from the absence of inhouse surgical support in the UK unit but both centres care for similar numbers of the target population i.e., <34-week infants. The	-

	Malaysian unit also cares for surgical infants, but these are in the majority term born infants with congenital malformations who are not included in this study. Both units have similar range of preterm infants admission per year and number of beds (25-26 units) and similar intensive medical care to the range of infants in this study.	
page 7, lines 11-15. The number of subjects should be based on power analysis. Was one done? If not, then state that this was simply a convenient number to collect.	No formal power calculation was conducted. The sample size was determined based on the usual admission numbers and length of stay at the neonatal units of both countries, with the aim of ensuring that daily data collection from birth to discharge was feasible for one person collecting data within the time available for the study. For both neonatal units, usual monthly admissions of preterm infants range from 30 to 50 infants. Therefore, collection of data from 50	This is explained on page 6.

	infants from each unit was deemed to be feasible and achievable.	
General: For "age" (as in weight for age) was gestational age used, or was it time since birth?	Corrected gestational age was used.	-
page 8, line 48 and other places. What tests were used? Please state here how you decided, and, when each p value is mentioned (in text or table) please state the test. In some cases, t-tests are clearly not right as the data are not close to normal.	P-values for comparisons between the two groups were determined by the Student's t-test or Mann-Whitney U test for continuous variables and by Chi squared or Fisher's exact tests for categorical variables, as appropriate.	Changes have been made on all table of results to indicate the tests used.
line 52-53. This is known as bivariate screening and it is not a good method. It can be shown that all the output from the multivariable equation is wrong (standard errors are too small, p values are too low, parameter estimates are biased away from 0; see Frank Harrell's book Regression Modeling Strategies for details). It's better to use substantive	Thank you for your suggestion. We have taken the approach used in many recently published studies in ADC F&N. Given the small number of observations at each site (n=50) this was considered an exploratory analysis only and so we limited the number of explanatory variables whose association with the change in Z-score we	-

knowledge, but if an automatic method must be used, LASSO isn't bad.	assessed. We included in the multivariable model both variables which were statistically significant in the univariable analyses, and those deemed to be clinically important based on established knowledge.	
line 58-60 Replace "correlation" with "collinearity" -- they are not the same.	Thank you. Changes have been made	Changes have been made on page 8
p. 10 line 35 and other places -- insert "significant" between "no" and "difference".	Thank you. Changes have been made	Changes have been made on page 10 and page 14
Table 3 - what is MD? I'm guessing it is mean difference, but please state this.	Yes, it is mean/median difference. Changes have been made, thank you.	Changes made on Table 3.
Reviewer 2		
The sample size could have been more.	We do agree, and we have added this as limitation in the paper. The sample size was determined based on the usual admission numbers and length of stay at the neonatal units of both countries, with the aim of ensuring that daily data	-

	collection from birth to discharge was feasible for one person collecting data in a relatively short period of time available for the study. For both neonatal units, usual monthly admissions of preterm infants range from 30 to 50 infants. Therefore, collection of data from 50 infants from each unit was deemed to be feasible and achievable.	
Reviewer 3		
What are the hospitals' criteria for discharge (which may have influenced length of stay)?	Both hospitals follow similar discharge criteria including weight of at least 1800g, not needing any additional medical support, and fully milk fed.	- Added this information on page 6
In estimating the nutritional content of breast milk, were gestational age and time postpartum taken into consideration?	This is described in the methods section for 'Baseline and feeding data' on page 7. We use the systematic review of preterm milk composition (Gidrewicz DA et al, 2014) as basis for our calculation which differentiate the	-

	calculation based on postpartum time.	
It is stated that “Postnatal growth failure was defined as a decrease in WAZ between birth and discharge of ≥ 1.28 as used in previous studies”. Both of the cited studies used a weight Z-score of <-1.28 at discharge as a criterion for growth failure, while Lin et al also use a change in Z-score of -1.28. The latter represents quite significant weight loss, while the former may represent a persistently small infant who tracked below the 10th percentile for weight. It would be helpful for readers to better understand the degree of weight loss (in grams or % birth weight) represented by this definition of growth failure.	In this study, infants < 34 weeks GA were included which mean that generalised weight loss calculation or comparison is not a suitable as main analysis for growth outcome due to heterogenicity of gestational ages, and degree of weight loss did not consider GA at discharge. As infants were generally being discharged home when about 1800g of weight is reached with satisfactory feeding progression and clinical conditions, degree of weight loss is expectedly minimal. The median degree of weight loss was at the peak for both units on week 1: UK at 5.7% and MAL at 4.4%.	We added the information on maximum degree of weight loss in Table 3.
Information about the method of assessment	Gestational age for both units were recorded in	-

and reliability of gestational age should be presented.	the electronic medical system: Caring Hospital Enterprise System (C-HEtS) in MAL unit and BadgerNet in the UK unit. The measurements were taken and checked/compared in both paper and electronic medical records to ensure consistency. As this is a prospective review of medical records, we do not have any access to check the reliability of this data. However, in general, both in the UK and Malaysia units, most infants' GA was determined by early first trimester ultrasound.	
Rare clinical conditions are noted as those that occurred in fewer than 10 babies per site, but this is 20% of the sample in each setting and thus not really rare.	We acknowledge that these are not really rare outcomes. For clinical conditions that were excluded from the univariable analyses, especially in the UK unit, only 1-6 infants were affected (2-12% of the sample). We chose not enter variables with these small number of infants in the model given the small sample size, the exploratory nature of the	We have removed the word 'rare' from the manuscript text and acknowledged that this decision was as a result of the small sample size.

	analysis, and the inability to assess effect sizes with any degree of precision.	
In table 2 values are show for the proportion of energy from breast milk and formula but does not indicate the time period this represents	This is the proportion analysed during whole admission. Changes have been made on Table 2	Changes have been made on Table 2
Details are needed on how cumulative energy and nutrient deficits or excesses were calculated. ESPGHAN reference values are presented in ranges: what values were actually used? Are these appropriate for the full range of gestational ages.	Deficits were calculated as the difference between the actual intake and the minimum intake recommended by the ESPGHAN recommendation. This specifies a minimum of 110 kcal/kg/d for energy intake, 3.5g/kg/d for protein (infants with >=1 kg birthweight), 4.0 g/kg/d protein (infants with <1 kg birthweight), 4.8 g/kg/d for fat, 11.6 g/kg/d for carbohydrate and 135 ml/kg/d for fluid.	Changes have been made on page 7 to include 'minimum' intake.
Table 4 should present sample sizes for each analysis	Thank you. I have added this info on Table 4.	Changes made on Table 4
Is there NO feeding of infants at the breast? Was all breast milk pumped and fed to	There was direct breastfeeding of the infants mostly during last few days before	We added the information on no record

infants? If not, how was intake estimated?	discharge home which was usually accompanied by bottle-feeding. Most infants were bottle-fed, and the volume was recorded as in this study. There was no record of volume of milk consumed via direct breastfeeding as we do not as a routine to record before and after feeding weight. This was also one of the reasons that average of intakes were done in clusters (week 1-4 or week 5-8) to accommodate for 'missing' volume of milk recorded especially in the last 2-3 days before discharge home.	of direct breastfeeding milk volumes on page 6.
In table 3, changes in weight Z-score should be adjusted for age duration of stay.	Changes in Weight-for-age Z-score is calculated based on infants gestational age at birth and at discharge. This has also been adjusted based on duration of stay in the multivariable regression model in Table 4.	-
In the statistical analysis section, it is stated that univariable analyses were used to explore	Yes. Given the small number of observations at each site (n=50) this was considered an	-

factors that predicted changes in WAZ". Do you mean bivariable (associations of each risk factor with the outcome, one at a time)?	exploratory analysis only and so we limited the number of explanatory variables whose association with the change in Z-score we assessed. We then entered variables significant in the univariable analyses into a multivariable model.	
The potential for selection bias related to missing data from earlier discharge should be more thoroughly discussed, since larger, healthier infants were likely discharged earlier and analyses such as those on nutritional intakes in weeks 1-4 will exclude the early discharges. Alternately, the analysis in table 4 could be restricted to intake in the first 2 weeks.	Median length of stay in this study was 32 days, in which 91% of data were available up until week 3 of admission and 62% of data were available at week 4. This was the reason why we use an average intake from week 1 to week 4 to calculate intakes and cumulative deficits which also helps to avoid potential multiple hypothesis testing with more weekly intakes comparisons.	We have expanded this in the discussion section: Page 12: "This may be due to smaller infants with greater burden of illnesses have longer stay and sub-optimal growth (27,28) while larger, healthier infants are more likely to be discharged earlier. "
Differences in nutritional intakes will result from the balance of feeding practices. The differences could be explained with analysis of the feeding data. Based	Thank you for your suggestion. We found and report the similarities and differences in feeding practices and have stated that the higher protein intake	-

on the results, it is clear that data on types and amounts of feeds are available. Thus, instead of speculating about the distinctions in practice, an important contribution would be to analyze how the different feeding strategies relate to nutritional intakes.	among Malaysian infants was due to more receiving PN and breast milk fortifier. We have added this further to the discussion. "The more frequent and longer use of PN and BMF resulted in differences in nutrient intakes particularly where Malaysian infants had higher protein intakes, lower cumulative deficits and earlier recovery of deficits. "	
Postpartum weight loss and regain is expected in all newborns. Time to regain birthweight is shown in table 3 but not discussed in the text. Moreover, other studies show that birthweight regain is related to mode of delivery: another variable that is shown in table 1 not discussed in the paper.	There were no significant differences between the two units in these variables. The data are provided in the table but not discussed in the text due to constraints of the word limit.	-
The objective of the study was to compare nutritional practices across neonatal units in very different settings,	We have discussed the differences in infants' characteristics being more SGA, more with lower birthweight and	Changes have been made on page 12.

and the paper presents a good description of the differences but offers little insight into what accounts for the differences. The second aim is to investigate the association of nutritional practices to postnatal growth. This requires a clear conceptual model. While the focus is on associations, not causal inference, there should nonetheless be some discussion of the direction of associations. A perspective that is lacking in the paper is that feeding practices are both a response to growth and other health conditions AND a predictor of growth. Change in feeding strategies may occur if an infant is not gaining weight. The discussion should acknowledge these important relationships and their implications for interpretation of the results.	more co-morbidities in MAL units that account for most differences in intakes especially higher number of infants on PN than in the UK unit. We agree that feeding practice in any neonatal units are a combination of following the feeding protocol based on infants' predictor of growth/demographic and a response to growth and other clinical conditions during admission. We have included this additional point in page 12.	

VERSION 2 – REVIEW

REVIEWER	Reviewer name: Dr. Peter Flom Institution and Country: Peter Flom Consulting, United States Competing interests: none
REVIEW RETURNED	14-Jul-2021

GENERAL COMMENTS	The authors have addressed my concerns and I now recommend publication.

VERSION 2 – AUTHOR RESPONSE

Comments	Responses and changes in main document
Associate Editor	
Abstract: In the design and methods, highlight this was an exploratory study See below for comments about the regression analysis and related conclusions, which may need to be modified here as well. If regression is included, the type of regression used should be given in the methods	Thank you for your suggestions. The exploratory nature of the study has been specified in the study title, objective and methods section.
What is known: Point three might be replaced with a point about the overall status of NICU science in the world - i.e, much is known about postnatal growth failure in HICs but in UMIC NICU science is rapidly expanding and robust and practice differences have not been well explored	Thank you for your suggestions. In response to this and further comments, we have revised this section (Page 4).
What this study adds: Given that this is an exploratory study with a very small sample, Im not sure any of these three comparative conclusions are justified. I recommend deleting and focusing more narrowly on the study's empiric	Thank you for your suggestions. We have included more specific findings in this section as per your suggestions. Changes have been made on page 4, as follows: 1. In this exploratory study, we found that infants

findings. What is clear is that the sample in Malaysia was more SGA, but this could just been random sampling given the small duration and N. You might say, cautiously, that in an exploratory study… these were the major clinical differences and practice differences between the sites. It also seems that there are differences in use of maternal milk which is very interesting and important but not highlighted in the manuscript. Apparently difference approaches to use of PN might also be important, but this is confounded by the fact that the Malaysia sample is so much more SGA	in the Malaysian unit were more likely to be small for gestational age and have co-morbidities. 2. Mother's own milk (MOM) is more frequently used in the Malaysian unit and more infants received supplementation with breast milk fortifiers. 3. More infants in the Malaysian unit received parenteral nutrition, although this may be a response to the greater nutritional needs of SGA infants.
Introduction/framing What the authors are doing here is very important, which is starting a conversation about comparisons between preterm infant characteristics and practices variation across different country income levels. It is true that most cross-site comparisons and registries have been just in HICs. But this point could be made more compellingly in the introduction, and that helps to justify better this small exploratory study which mostly proves the feasibility of doing such comparisons, as the actual noted differences are difficult to interpret given the small sample size and lower birth weight in Malaysia (which might just be due to chance). Also in the introduction and throughout make the note that Malaysia is an UMIC, which is a very	Thank you for these comments. We have re-written the introduction in keeping with your suggestions (please see page 5). This includes: 1. Highlighted the lack of studies comparing HIC with MIC/UMIC: "not much research has explored the impact of improvement in neonatal care or considered whether it is feasible to compare nutritional practices in upper-middle income settings with a higher income country setting. " 2. Corrected that Malaysia is an UMIC and not MIC

specific comparison that might not be applicable to lower income countries.	
Methods/Results/Discussion Emphasize in the methods section on sample size that this was exploratory, and the n was chosen based on time and personnel constraints alone	Thank you for your suggestions. We have added this information. Changes has been made on page 6, as follows: “As this was an exploratory study, the sample size was determined based on the usual monthly admissions and length of stay at the respective units. Therefore, collection of data from 50 infants from each unit (total of 100) was deemed to be feasible within the time and resources available for the study.”
In the first paragraph mention is made of quantile regression, but then the description of the regression method it is not clear if this is quantile regression. It should be made more clear what the regression model is- quantile regression for continuous change in WAZ? This can be more clear in the results section and table also. Mention is made in several places of the proportion of infants meeting the -1.28 Z score definition of postnatal growth failure, and this may generate confusion for readers in terms of what is being regressed. Overall (see below) I don’t think the regression models help	Quantile regression was not used in this study. In keeping with this and further comments, we have removed the section on regression analysis.
Overall in this revised version I still question the utlility of the regression analysis in Table 4. The two sites are very different based on the proportion of SGA and overall unwellness. But the paper as such doesn’t provide us with any details sociodemographic or	Thank you. We have removed the regression analysis. We have included more information in Results and Discussion based on Tables 2 and 3.

clinical data that might help explain this difference in proportion of SGA (which might be due to chance given the fact that the sample is so small and not systematic). And the variables presented in Table 4 might be predictors of poor nutrition OR clinical responses to poor nutrition (starting PN, increasng protein intake etc) as one of the reviewers in the first round notes. Overall I think the regression analysis detracts from Table 3 which is much more interesting. The use of parsimonious variable reduction makes it even less useful, since at the end of the day the final regression model for Malaysia just shows that being SGA leads to a long hospital stay. I would suggest dropping the regressions and focusing more on the empirical differences, such as maternal breast milk usage.	We agree that the lack of maternal demographic information is a limitation of the study. We have added this to the discussion and have written about this in the limitation section “Maternal details including her nutritional status, clinical conditions, and antenatal care should also be considered as these are vital determinants of fetal nutrition and infants’ outcomes including SGA status “. We have also stressed in the conclusion that future studies must include this. Table 4 has been removed to avoid such misinterpretation of our findings.
The discussion would be more effective if focused on clear empirical differences. Most of the discussion of differences observed in the WAZ discharge criteria are not justified given the small and unrepresentativeness of the sample in both countries, and the lack of detailed clinical information. For example it isn’t useful to speculate about genetic differences or LBW rates (which is not the same thing as SGA) in Malaysia based on a sample of 50 infants.	Thank you for your suggestions. The discussion is re-written with focus on the empirical differences and removal of the regression analysis.
If the regression analysis is removed (my recommendation) or greatly deemphasized (if it is left in I would not show the adjusted models, given	Thank you for your suggestions. We have removed the regression analysis and used the word count to discuss the empirical differences as you suggest.

the low N and asymmetry between the two models, just the bivariate analysis), there is also more space to discuss what is truly novel here, which is the feasibiity of comparing NICU practices across country income levels. The discussion should also building on the feasibility work and describe what next steps look like - how do we begin generating systematic comparative outcome data across country/income levels?	We have also stated that this exploratory work is evidence that larger studies are feasible and needed with suggestion for design of such studies. "Studies with an adequately powered sample, collection of more data on maternal characteristics and infants' longer term outcomes and the use of a more representative growth chart would provide evidence to ensure that preterm infants receive adequate nutrition, hopefully, in all care settings globally. "
Conclusions Should focus on the feasibility work and next steps. The current conclusions are not justified, as they make inferences from the regression models which are undermined by the very significant clinical characteristics of infants in the two sites.	Thank you for your suggestions. The conclusion has been restated as: "In our exploratory analyses, there were variations in nutritional practices between the two units included in the study. Current nutritional practices often do not meet recommended intakes, especially for protein in preterm infants. We found that with international collaborations, future comparison studies involving units in varied income settings are feasible and may provide evidence to support equity in care of preterm infants."
"maximum fluid intake in both the cohorts seem less than the recommendation (160-180ml/kg) " - comment on why fluid intake was at this level (lower than recommendation) in both cohorts	In this study, we based our analysis by using ESPGHAN 2010 recommendation which suggests the range of 135-200 ml/kg/d for fluid. In our experience, these rates are not unusual and hence we did not comment on this. Full enteral feed is generally considered as 140-150 ml/kg/day in preterm infants in the UK. While many will receive higher volumes if needed, especially when on unfortified expressed breast milk, the volumes achieved in the study infants

	are well within our practice recommendations. In addition, more infants in the Malaysian unit were challenged with higher incidence of co-morbidities such as chronic lung disease and patent ductus arteriosus that likely necessitated the restriction of total fluid intake.
"Discussion, second paragraph: "the enteral feeding initial=n and progression are no different. is it possibly due to the local PN protocols?" - most of this difference is just due to SGA proportion. This should be clearer.	Thank you. Please see if the changes made to the discussion section have resolved this.
"Information about the method of assessment and reliability of gestational age should be presented." - please provide details in the manuscript on local methods for determining GA	Thank you. We have added this: "In the Malaysian unit, GA is determined by using early first trimester ultrasound or by estimation based on last menstrual period for those who presented in later pregnancy. In the UK unit, GA was determined by early first trimester ultrasound. These records were retrieved from both paper and electronic medical records."
"Rare clinical conditions are noted as those that occurred in fewer than 10 babies per site" - the term rare is still used in the manuscript	We have removed this mention.
EiC (acting, as prof Choonara has a COI)	
Is the first bullet point of the what this study adds really novel ? Bullet point 3, new: what is covered with 'demographic' differences, does this include differences in clinical characteristics between both cohorts ?	This section has been re-written

is the wording upper, middle and high incomes correct (reads different from the introduction, you aimed to compare a low and a high income setting ?	Malaysia is an upper-middle income country, we have corrected this at each mention.
Introduction, 'can be applied' = is 'can' the best wording, or does this also relates to other circumstances (in the next sentence, you refer to this aspect)	We hope this has been resolved with the re-writing of the introduction.
Methods Do I understand this correct that the 'power' calculation was rather based on feasibility, and not to explore potential differences ? If so, table 3 is rather explorative, but not the powered primary outcome ?	Yes – the limitation of the lack of power calculation has been clarified further as mentioned above. The analyses are exploratory. We hope that this is clearer now that we have specified it in various sections. We have therefore removed Table 4 and the regression analysis to avoid any misinterpretation of the results in the absence of adequate power.
How has feeding been handled in the event of 'direct' breastfeeding ? omission is likely not correct ?	There was direct breastfeeding of the infants mostly during last few days before discharge home which was usually accompanied by bottle-feeding. Where infants were bottle-fed (formula or expressed breast milk) the volume was recorded. There was no record of volume of milk consumed via direct breastfeeding. This was one of the reasons that average of intakes were done in clusters (week 1-4 or week 5-8) to accommodate for 'missing' volume of milk recorded especially in the last 2-3 days before discharge home.
and how has (product, %) fortification been handled.	Amount of fortifier added to specified volumes of milk were recorded and produce specifications were used to calculate the resulting intakes. Results in Table 3 showed proportion

	of infants who received any BMF during admission.
It is reasonable to use the Fenton growth chart for both UK and Malaysian infants ? please elaborate on this.	We have written this in the methods section. We have used the same chart on both populations to enable a comparison. Fenton growth chart is the standard reference growth chart use in many countries including Malaysia, although the basis of data used for the construction of this chart comprises data collection from mostly Caucasian population from Germany, United States, Canada, Australia, Scotland and Italy. However, this chart is used in Malaysian unit as it also links to the WHO growth data from birth up until 10 weeks post-term. To ensure consistency in attribution of SGA status, determination of Z-scores, and assessment of Z-score change over time between these two study sites, Fenton was chosen as its data were based on large number of samples, and with more recent data as compared to other chart such as UK-WHO's. The recent INTERGROWTH-21st can also be used, but it does not have enough data prior to 33 weeks, and also have only small numbers at 33 to 34 weeks making it to only be a suitable tool for monitoring the growth of preterm infants who are born at ≥32 weeks' gestation up to 6 months' post term-corrected age.
There was a higher % of SGA cases, but is this due to the Fenton chart, or are there other reasons, like maternal morbidity characteristics ?	We agree that it is possible that % SGA is higher in Malaysian cohort due to the use of Fenton chart which based on Caucasian population. However, this chart is routinely used in the unit and was therefore selected for the comparison.

	Maternal characteristics and other perinatal factors are likely to be the cause of the higher SGA rates. Unfortunately, we do not have the data to explore this. We have written this as a limitation and explained further in the discussion section.
As the study was prospective, are there other 'time points' on weight Z scores besides on admission and at discharge ? You suggest this as a strength, but do not really explore or report these data.	Thank you. We have weekly measurements of weight Z-score from postnatal week 1 until week 8/discharge. This has been added on Page 11 and Figure 1
Ethics: EC approval for a prospective study without any information to parents is at least in my setting not possible, but I accept this construct, but highly recommend to add that this approach was accepted by the EC.	As the study was entirely observational and involved collection of routinely recorded clinical data, we were granted ethical approval in keeping with UK and Malaysian regulations. We have added this in the Ethical approval section that this approach was accepted by EC in both countries. Changes has been made on page 8
Results: why is the use of BMF so different between both cohorts (assuming that the 26 % refers to the MOM cases only), and does this 'fully' explains the differences in protein vs caloric intake (cf section nutritional intakes).	The two units follow different protocols for use of BMF. This and the higher breastfeeding rate in the Malaysian unit explains the greater use of BMF in that unit. The higher protein intake is due to greater use of BMF and more frequent and prolonged use of PN. This is also explained in more details in the discussion section. We have described this in the discussion: "Supplementation of breast milk with BMF or protein supplements was more frequent in the Malaysia unit. BMF was used more selectively in the UK unit. The standard protocol recommends the addition of BMF at 75-100ml/kg/d milk feeds in Malaysian unit

	while in the UK unit, protocol suggests the addition of BMF when feeding reaches 150-180 ml/kg/ but only at clinician's discretion when there are significant concerns about growth. Interestingly, the majority of infants who received breast milk in the Malaysian unit received some fortification, while the majority of infants who received breast milk in the UK unit received supplemental formula feeding. Interestingly, in the Malaysian unit, infants who were mixed fed i.e. had breast and formula milk, continued with breast milk fortification while most mixed fed infants in the UK unit did not receive BMF. "
In the discussion, you suggest 'selective' use of BMF in the UK unit, but how was selection done ?	Thank you – we have clarified this in the discussion. We have added the following to the discussion: "The standard protocol recommends the addition of BMF at 75-100ml/kg/d milk feeds in Malaysian unit while in the UK unit, protocol suggests the addition of BMF when feeding reaches 150-180 ml/kg/ but only at clinician's discretion when there are significant concerns about growth."
I may have missed this, but how has PN been defined (? Protein ? protein and fat ?)	We analysed PN in this study based on its protein, lipid and carbohydrate contents (glucose %). We have clarified this in the methods section "The nutritional content of EN and PN (protein, lipid, and carbohydrate content) were calculated based on the manufacturers' literature while the composition of breast milk was based on current evidence"
Table 1: I recommend to verify this table, as you likely report IQR and range for ? gestational age, ? similar	Thank you for your feedbacks. We have edited for IQR data only on Table 1.

for HC-for-age Z score ? parity ? CLD is on day 28, or week 36 ?	-Definition of CLD has been added to the footnote.
diagnosis ROP, ICH, PDA: suggest to explain how this has been handled (eg pda screening, or in selected cases, or treated cases ?)	All definitions have been added to the footnote.