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.

ARTICLE DETAILS

TITLE (PROVISIONAL)	A prospective observational study of Universal Newborn Eye Screening in a hospital and community setting in New Zealand
AUTHORS	Simkin, Samantha; Misra, Stuti; Battin, Malcolm; McGhee, Charles; Dai, Shuan

VERSION 1 – REVIEW

REVIEWER	Reviewer name: Dr Sarah J Nevitt Institution and Country: University of Liverpool United Kingdom Competing interests: I have no competing interests
REVIEW RETURNED	16-Oct-2018

GENERAL COMMENTS	I have performed a statistical review of the manuscript “New Zealand prospective Universal Newborn Eye Screening reveals frequent retinal haemorrhages and ocular abnormalities.” The authors describe a prospective cohort study of Universal Newborn Eye Screening and statistical analysis of factors related to retinal haemorrhages. I have several comments on this work for the attention of the authors: 1) Objective of the work (i.e. the role of UNES) Page 6, line 36: “Infants with RH on initial screening were re-screened at six weeks to assess the status of retinal haemorrhages...” I don’t have much clinical knowledge in this area, so just to clarify, any RH observed on the initial screening would be observed and RH would either resolve spontaneously (without treatment?) or remain at six weeks? Could RH or any other abnormality get worse after six weeks follow-up (with or without treatment)? Also, were only the infants with RH present on the initial screening invited back for follow-up or all infants? In other words, if RH was absent at the first screening, it would not develop at a later stage? The point of my comment here is that I’m not completely following what the benefits are of UNES compared to the usual screening (red reflex test?) Is the idea to identify and treat abnormalities earlier? Or if these abnormalities are not treated and/or resolve spontaneously, what is the benefit to identifying earlier or identifying more abnormalities 2) Demographic information Table 1: I understand the rationale of this table (i.e. to demonstrate how applicable this sample of infants is to the regional newborn population to generalise findings).

	However, I'm not sure how useful it is to present comparative information on different levels (i.e. infant level delivery modality and ethnicity for this study and mother level delivery modality and ethnicity for the regional data). It would be preferable to present data on the same levels where possible. I'm also not following why delivery modality (CS and SVD) are highlighted specifically as being higher or lower than regional data. Were these variables statistically tested? And if so, were the differences in the level of data (i.e. infant vs mother level variables) taken into account? 3) Unit of analysis Related to the above comment regarding infant and mother level variables in Table 1, I am finding it a little difficult to follow the level / unit of analysis across the manuscript – i.e. which characteristics are measured at the infant level, mother level etc. within the present study and within the comparative data. Furthermore, I understand that RH can be unilateral (one eye) or bilateral (both eyes) so do I assume correctly that RH would be classified as present in an infant if it is present in one or both eyes rather than eye being the unit of analysis when calculating prevalence? i.e. the 50 cases refers to 50 infants with RH present rather than 50 eyes with RH present? Could the authors please also clarify whether multiple births (i.e. twins, triplets etc. from the same birth) were included in the study? If this is the case then then independence across the characteristics considered in Table 2 cannot be assumed (i.e. multiple children from the same birth would have the same delivery modality presumably, and the same GA, the same maternal height etc.). The statistical methods used assume independence so if infants from multiple births are included then the statistical tests are not valid. Also, I'm not aware if any of the abnormalities of interest could be considered genetic? If so, could the authors confirm if any siblings or other genetically related infants were permitted within the study? If so, as above independence cannot be assumed as the statistical tests would not be valid. 4) Statistical testing If all infants can be assumed to be independent (see comment 3), please clarify which test is used to calculate the p value for GA (median) in Table 2 Please also clarify for the regression whether the odds ratios presented are from Univariable regressions (i.e. each characteristic entered into a regression model as a single variable) or from multivariable regression (i.e. all characteristics entered into a regression model together). Also were regressions performed for only significant p values (from Table 2)? For completeness, I suggest reporting Odds ratios for all characteristics regardless of statistical significance. 5) Interpretation The author's comments on "When compared to caesarean section, logistic regression identified a 26.30 times increase in the odds of RH in infants born by normal vaginal delivery and a 33.61 times increase with instrumental vaginal delivery…

	” I suggest that more context should be given to these results. The reference category (CS) had very few events, hence these regression results are associated with a large amount of imprecision – this should also be acknowledged within the interpretation 6) What this study adds Related to comment 1 (if only those with RH are invited back for follow-up), I’m not sure that this data can be considered ‘longitudinal.’ Also, there is no data past six weeks of age so persistence after 6 weeks is unknown. I suggest rewording the last point. The authors note that previous studies have considered only hospital settings (page 4), therefore this study considering both hospital and community settings could be considered novel. Therefore did the authors consider including hospital vs community setting in their analyses (Table 2) to examine any differences? I also have a few minor comments, mainly on wording: Page 4, line 23: “This screening method has been incorporated into many retinopathy of prematurity screening programs worldwide, significant incidental findings of ocular abnormalities reported.” I think there is a word missing from this sentence somewhere? Page 4, line 32: “Prevalence of ocular abnormalities ranged from 24.4% and 4.7%” Quoting prevalence from high to low is unusual and other figures in the manuscript are quoted from low to high (e.g. within the discussion, page 11). I suggest being consistent throughout the manuscript Page 5, line 16: “Infants who were already enrolled in ROP screening…” Please define ROP Page 10, line 48: “UNES detected ocular abnormality in 54 out of 346” The results section implies 55 (50 RH (page 8, line 38) plus other ocular abnormalities in 5 infants (page 10, line 23). Did one child have both RH and another abnormality? Please clarify Page 11, line 3: The authors note the poor sensitivity of the red reflex test. However this study does not measure sensitivity and specificity of the UNES so this does not seem a fair comparison at this point – further context of this comment regarding sensitivity is needed. Perhaps consideration of the sensitivity and specificity of UNES could be a recommendation from the authors for further research (page 13, line 44)?

REVIEWER	Reviewer name: Catherine cassiman Institution and Country: UZ Hospital Leuven Competing interests: pediatric ophthalmology, strabismus, neuro-ophthalmology
REVIEW RETURNED	17-Oct-2018

GENERAL COMMENTS	interesting study; comprehensively presented; lack some ophthalmological information

REVIEWER	Reviewer name: Birgit Lorenz, MD, PhD, FEBO, FARVO Institution and Country: Dept of Ophthalmology, Justus-Liebig-University Giessen, Friedrichstrasse 18, 35390 Giessen<, Germany Competing interests: no competing interests
REVIEW RETURNED	11-Nov-2018

GENERAL COMMENTS	The study as such is interesting but not really novel besides the fact that it is from another continent i.e. New Zealand, compared to previous reports from China and India. Similar to previous reports the high percentage of pathologies are mostly transitory retinal hemorrhages that usually remain without functional sequelae. The comparison of 2 settings i.e. hospital born babies and babies born outside hospitals reduces the power of the study as the numbers of serious pathologies are quite small although early detection especially of cataract is desirable. The cataract however is also taken up by the red reflex examination. Optic nerve hypoplasia manifests by the appearance of infantile nystagmus from about 8 weeks on. To have robust numbers with regard to the rare pathologies a much higher number of babies would have to be examined (likely at least 5000). Apart from this limitation the paper is well written, but does not add much to the current literature.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1:

I have performed a statistical review of the manuscript “New Zealand prospective Universal Newborn Eye Screening reveals frequent retinal haemorrhages and ocular abnormalities.” The authors describe a prospective cohort study of Universal Newborn Eye Screening and statistical analysis of factors related to retinal haemorrhages.

I have several comments on this work for the attention of the authors:

Comment 1: Objective of the work (i.e. the role of UNES) Page 6, line 36: “Infants with RH on initial screening were re-screened at six weeks to assess the status of retinal haemorrhages...”

I don’t have much clinical knowledge in this area, so just to clarify, any RH observed on the initial screening would be observed and RH would either resolve spontaneously (without treatment?) or remain at six weeks? Could RH or any other abnormality get worse after six weeks follow-up (with or without treatment)? Also, were only the infants with RH present on the initial screening invited back for follow-up or all infants? In other words, if RH was absent at the first screening, it would not develop at a later stage?

The point of my comment here is that I’m not completely following what the benefits are of UNES compared to the usual screening (red reflex test?)

Is the idea to identify and treat abnormalities earlier? Or if these abnormalities are not treated and/or resolve spontaneously, what is the benefit to identifying earlier or identifying more abnormalities?

Response 1: Thank you for your comments. Red reflex has very poor sensitivity for posterior eye disease (4%), therefore the benefit of UNES is to assess the newborn eye in more detail for early detection of abnormalities, leading to early treatment and intervention where appropriate.

Birth-related retinal haemorrhages are expected to resolve by six weeks of age, and if not present at first examination would not occur at a later date as they are thought to be due to the birth process. After the age of six weeks it is important to have clear vision for normal neurological visual development, therefore observation for resolution is important. As we were looking for abnormalities which were present from birth infants without retinal haemorrhages were not reviewed at six weeks, although any child with any other ocular abnormality detected at birth were appropriately referred for management. Ocular abnormalities detected at birth, other than birth-related retinal haemorrhages are not expected to resolve spontaneously and therefore early intervention is essential. To clarify this for readers we have added a small phrase at the end of the seventh paragraph in the methods '…as other ocular abnormalities detected at birth, other than birth-related retinal haemorrhages are not expected to resolve spontaneously.'

Comment 2: Demographic information

Table 1: I understand the rationale of this table (i.e. to demonstrate how applicable this sample of infants is to the regional newborn population to generalise findings). However, I'm not sure how useful it is to present comparative information on different levels (i.e. infant level delivery modality and ethnicity for this study and mother level delivery modality and ethnicity for the regional data). It would be preferable to present data on the same levels where possible.

I'm also not following why delivery modality (CS and SVD) are highlighted specifically as being higher or lower than regional data. Were these variables statistically tested? And if so, were the differences in the level of data (i.e. infant vs mother level variables) taken into account?

Response 2: Thank you for your comment Reviewer 1, we agree that it would be desired to present this data on the same levels where possible. Unfortunately, this hospital demographic data is reported at a maternal level so we cannot do a formal comparison. However, we considered it necessary to indicate the representativeness of our cohort. Modality differences were highlighted mainly to describe our cohort, and that caesarean sections may have been over represented in our cohort due to their longer stays in hospital.

Comment 3: Unit of analysis

Related to the above comment regarding infant and mother level variables in Table 1, I am finding it a little difficult to follow the level / unit of analysis across the manuscript – i.e. which characteristics are measured at the infant level, mother level etc. within the present study and within the comparative data. Furthermore, I understand that RH can be unilateral (one eye) or bilateral (both eyes) so do I assume correctly that RH would be classified as present in an infant if it is present in one or both eyes rather than eye being the unit of analysis when calculating prevalence? I.e. the 50 cases refers to 50 infants with RH present rather than 50 eyes with RH present?

Could the authors please also clarify whether multiple births (i.e. twins, triplets etc. from the same birth) were included in the study? If this is the case then then independence across the characteristics considered in Table 2 cannot be assumed (i.e. multiple children from the same birth would have the same delivery modality presumably, and the same GA, the same maternal height etc.). The statistical methods used assume independence so if infants from multiple births are included then the statistical tests are not valid.

Also, I'm not aware if any of the abnormalities of interest could be considered genetic? If so, could the authors confirm if any siblings or other genetically related infants were permitted within the study? If so, as above independence cannot be assumed as the statistical tests would not be valid.

Response 3: We have reported all of our findings at the infant level throughout the manuscript, the manuscript has been reviewed to confirm the consistency of this use and clarify any queries. Thus should be easier for the reader to follow. Specifically results are reported per infant rather than per eye, i.e. there are 50 children with retinal haemorrhages, who were recorded to have retinal haemorrhages whether or not this was bilateral or unilateral.

There were 11 sets of twins in this study and no siblings. Analysis of the data was done with the support of an experienced biostatistician. Due to the small number of twins and the majority of abnormalities to be detected not being genetic it was decided to leave all infants in the analysis and statistically analyse them as though they were independent. We have added a sentence to the results stating this to prevent ambiguity for the reader. This sentence reads 'There were 11 sets of twins in the study, all were included in the analysis. Although outcomes of twins could potentially be genetically linked, we did not observe this in our study with the only abnormalities detected in twins being RH.'

Comment 4: Statistical testing

If all infants can be assumed to be independent (see comment 3), please clarify which test is used to calculate the p value for GA (median) in Table 2

Please also clarify for the regression whether the odds ratios presented are from Univariable regressions (i.e. each characteristic entered into a regression model as a single variable) or from multivariable regression (i.e. all characteristics entered into a regression model together). Also were regressions performed for only significant p values (from Table 2)? For completeness, I suggest reporting Odds ratios for all characteristics regardless of statistical significance.

Response 4: Thank you for your thorough statistical review of our work, as stated in response 3 we worked alongside an experienced biostatistician for our data analysis. Difference in gestational age was analysed using the Mann-Whitney U Test, the U = 6072, and the p value = 0.042, I have included the U-value in the Table 2 for completeness of reporting. The analysis was univariable and as recommended by the statistician only statistically significant variables were included in this.

Comment 5: Interpretation

The author's comments on "When compared to caesarean section, logistic regression identified a 26.30 times increase in the odds of RH in infants born by normal vaginal delivery and a 33.61 times increase with instrumental vaginal delivery…" I suggest that more context should be given to these results. The reference category (CS) had very few events, hence these regression results are associated with a large amount of imprecision – this should also be acknowledged within the interpretation

Response 5: Thank you for highlighting this, we have added a sentence to the discussion that states 'There is potential for imprecision in the odds ratio calculation due to the limited number of events in the caesarean section category, hence larger studies would be beneficial.'

Comment 6: What this study adds

Related to comment 1 (if only those with RH are invited back for follow-up), I'm not sure that this data can be considered 'longitudinal.' Also, there is no data past six weeks of age so persistence after 6 weeks is unknown. I suggest rewording the last point.

Response 6: We have reworded this last point as suggested and to prevent confusion have simply removed the word 'longitudinal', in addition we have added that the majority of retinal haemorrhages resolve spontaneously (see Editor-in-Chief comment/response 9).

This sentence now reads ‘Data on retinal haemorrhages among infants screened indicate that some retinal haemorrhages persist past six weeks of age, with unknown long term consequences, but the majority resolve spontaneously.’

Comment 7: The authors note that previous studies have considered only hospital settings (page 4), therefore this study considering both hospital and community settings could be considered novel. Therefore did the authors consider including hospital vs community setting in their analyses (Table 2) to examine any differences?

Response 7: In Table 2 the authors compared variables with potential biologic plausibility of having an effect on RH. The importance of hospital vs community is the potential to utilise the programme in a ‘universal’ manner. We have included some analysis of the % of haemorrhages in both the hospital and community settings with the following sentence added to the results section, ‘Infants screened in the community setting were more commonly screened before 72 hours ($p<0.001$), this reflect in the percentage of haemorrhages present with 25% of infants in the community setting and 8.3% in the hospital ($p<0.001$).’

This has also been highlighted in the discussion, ‘There was a statistically significant difference in the rate of haemorrhages present in community compared to the hospital which aligned with the significantly earlier screening occurring in the community.’

Comment 8: Page 4, line 23: “This screening method has been incorporated into many retinopathy of prematurity screening programs worldwide, significant incidental findings of ocular abnormalities reported.” I think there is a word missing from this sentence somewhere?

Response 8: Thank you for your observation, we have amended the sentence and it now reads, “This screening method has been incorporated into many retinopathy of prematurity screening programs worldwide with significant incidental findings of ocular abnormalities reported.”

Comment 9: Page 4, line 32: “Prevalence of ocular abnormalities ranged from 24.4% and 4.7%”

Quoting prevalence from high to low is unusual and other figures in the manuscript are quoted from low to high (e.g. within the discussion, page 11). I suggest being consistent throughout the manuscript

Response 9: The order of the prevalence range reported has been from lowest to highest and have ensured this is consistently used throughout the manuscript.

Comment 10: Page 5, line 16: “Infants who were already enrolled in ROP screening…” Please define ROP

Response 10: This sentenced has been amended with retinopathy of prematurity (ROP) no longer abbreviated, for clarity the sentence now reads, ‘Infants who were already enrolled in retinopathy of prematurity screening (infants born <30 weeks gestational age or <1250 gram birth weight) were excluded from this study.’

Comment 11: Page 10, line 48: “UNES detected ocular abnormality in 54 out of 346”

The results section implies 55 (50 RH (page 8, line 38) plus other ocular abnormalities in 5 infants (page 10, line 23). Did one child have both RH and another abnormality? Please clarify

Response 11: Thank you for highlighting the lack of clarity in this sentence, yes one child did have both an RH and another abnormality. To ensure clarity for readers we have amended the sentence to, ‘UNES detected ocular abnormality in 54 of the 346 infants who completed screening, 14.5% of participants had RH and 1.4% had other ocular abnormalities, with one child having both RH and another ocular abnormality.’

Comment 12: Page 11, line 3: The authors note the poor sensitivity of the red reflex test. However this study does not measure sensitivity and specificity of the UNES so this does not seem a fair comparison at this point – further context of this comment regarding sensitivity is needed. Perhaps consideration of the sensitivity and specificity of UNES could be a recommendation from the authors for further research (page 13, line 44)?

Response 12: Thank you for highlighted our goals for future research, we have amended the manuscript to now read: “Ocular abnormality screening at birth is with the “red reflex test” in many countries, including New Zealand.1,2 However, this misses a significant proportion of posterior ocular abnormalities with a sensitivity of only 4%,5 This indicates the potential need for other newborn ocular screening methods.” To clarify that we are not comparing sensitivity and specificity between these two methods, but rather exploring new options for screening. In addition we have amended the conclusion to state ‘In addition, a full economic analysis, sensitivity and specificity calculations, as well as implementation feasibility must be determined prior to initiation of any large-scale screening.’

Reviewer: 2

Comment 1:

interesting study; comprehensively presented; lack some ophthalmological information on how hypoplasia was diagnosed.

Response 1: Thank you for your positive review, a sub-specialist paediatric ophthalmologist reviewed all images and determined if an abnormality was present, in the case of optic nerve hypoplasia this was based on disc size relative to the disc to macula size. Infants with an abnormality identified at screening were referred for clinical management and diagnosis, both infants with optic nerve hypoplasia had this diagnosis confirmed.

Reviewer: 3

Comments to the Author

Comment 1: The study as such is interesting but not really novel besides the fact that it is from another continent i.e. New Zealand, compared to previous reports from China and India. Similar to previous reports the high percentage of pathologies are mostly transitory retinal haemorrhages that usually remain without functional sequelae. The comparison of 2 settings i.e. hospital born babies and babies born outside hospitals reduces the power of the study as the numbers of serious pathologies are quite small although early detection especially of cataract is desirable. The cataract however is also taken up by the red reflex examination. Optic nerve hypoplasia manifests by the appearance of infantile nystagmus from about 8 weeks on.

To have robust numbers with regard to the rare pathologies a much higher number of babies would have to be examined (likely at least 5000).

Apart from this limitation the paper is well written, but does not add much to the current literature.

Response 1: Thank you for your review of our study, which adds to the body of evidence for universal newborn eye screening. We agree that large numbers are needed for those rare pathologies.