PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (see an example) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

This paper was submitted to the ADC but declined for publication following peer review. The authors addressed the reviewers' comments and submitted the revised paper to BMJ Open. The paper was subsequently accepted for publication at BMJ Open.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Respiratory infections in preterm infants and subsequent asthma
AUTHORS	Montgomery, Scott; Bahmanyar, Shahram; Brus, Ole; Hussein, Oula; Kosma, Paraskevi; Palme-Kilander, Charlotte

VERSION 1 - REVIEW

REVIEWER	***The reviewer wished to be anonymous.***
REVIEW RETURNED	16-Aug-2013

GENERAL COMMENTS	This study considers the association between airway infections in the 1st year of life and asthma after age 5. How gestational age may modify this association is also investigated. The cohort used for the study appears appropriate and the models fitted reasonable. However, I do have some queries and comments about the analysis that need to be addressed before publication: 1. Controls are selected as matches for the cases – up to 5 per case, yet this matching is subsequently ignored. It is usual to retain matching in the analyses to increase efficiency. Why was this not done? What effect might this omission have on the results? Should matched analyses replace those given? 2. Numeric variables are categorised and then modelled. Such categorisations are not generally recommended. - Models based on these will be less powerful than those using continuums and cannot lead to a clinically plausible model (risk will increase gradually as, for example, gestational age or brithweight decreases, it does not suddenly change at some cutpoint). 3. 'Multivariate' is used where 'multivariable' is meant. 4. Please state how the proportional hazards assumption is tested and give results of this investigation. 5. Stratified analyses are given in table 4. Please clarify how these models have been fitted. In particular, are these models each based on separate subsets of the data (within strata) and hence differing adjustments made for potential confounders such as sex, or are the confounders constant in effect across strata (ie. a combined model is fitted with differing coefficients allowed for infection association only within gestational age groups)? If it is not the latter, why not? It does not seem reasonable to allow all coefficients to vary at each gestational age group.

	6. It is unclear how exactly the interaction p-value was obtained. I would have expected a single gestational age by infection interaction term given with confidence intervals, but this is assuming gestational age as a continuum. Please clarify the investigation of interaction and give more details of effect size with confidence interval. As mentioned in point 2 above, I am not keen on the categorisation of gestational age and would either expect to see gestation modelled as a continuum (plus interactions) or full justification of why this was not done.

- The manuscript received two reviews at The ADC but the reviewers have declined to make the reviews public. Please contact BMJ Open editorial office for any further information.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Comments to the Author
This study considers the association between airway infections in the 1st year of life and asthma after age 5. How gestational age may modify this association is also investigated. The cohort used for the study appears appropriate and the models fitted reasonable. However, I do have some queries and comments about the analysis that need to be addressed before publication:

1. Controls are selected as matches for the cases – up to 5 per case, yet this matching is subsequently ignored. It is usual to retain matching in the analyses to increase efficiency. Why was this not done? What effect might this omission have on the results? Should matched analyses replace those given?

RESPONSE: The analysis includes interaction testing using one of the characteristics (gestational age) used for matching the exposed and unexposed cohorts; therefore it would not be possible to perform this interaction testing if the models were internally stratified by the matching characteristics (the non-interaction results are almost identical). We have extended the discussion section on this topic as follows: As interaction testing by gestational age was undertaken, adjustment rather than internal stratification for this factor was performed.

2. Numeric variables are categorised and then modelled. Such categorisations are not generally recommended. - Models based on these will be less powerful than those using continuums and cannot lead to a clinically plausible model (risk will increase gradually as, for example, gestational age or brithweight decreases, it does not suddenly change at some cutpoint).

RESPONSE: We chose to use categories for some measures, as we did not want to assume that the associations were linear, as appears to be the case for gestational age. We used the recognised cut-points for gestational age, which also helps with the clinical interpretation of the paper's findings. Birth weight was always modelled as a continuous variable when used for adjustment in multivariable analysis (see statistics section of the original manuscript and the augmented text in version two). It was presented categorically in table 2 purely for descriptive purposes to demonstrate variation (which was very slight) between the exposed and unexposed cohorts.

3. 'Multivariate' is used where 'multivariable' is meant.

RESPONSE: This has been corrected (in the statistical analysis section).

4. Please state how the proportional hazards assumption is tested and give results of this

investigation.

RESPONSE: Due to the relatively large sample size, log-minus-log plots were used to assess whether the proportional hazards assumption was violated: the curves comparing the exposed and unexposed cohorts did not converge and proportionality was maintained. This information has been added to the statistical analysis section of the paper.

5. Stratified analyses are given in table 4. Please clarify how these models have been fitted. In particular, are these models each based on separate subsets of the data (within strata) and hence differing adjustments made for potential confounders such as sex, or are the confounders constant in effect across strata (ie. a combined model is fitted with differing coefficients allowed for infection association only within gestational age groups)? If it is not the latter, why not? It does not seem reasonable to allow all coefficients to vary at each gestational age group.

RESPONSE: For the interaction analysis, a single model was used. The stratified analyses are also useful for more straightforward interpretation by less statistically inclined readers. The statistical analysis section has been altered to make it clearer that the main results investigating interactions used the entire dataset together.

6. It is unclear how exactly the interaction p-value was obtained. I would have expected a single gestational age by infection interaction term given with confidence intervals, but this is assuming gestational age as a continuum. Please clarify the investigation of interaction and give more details of effect size with confidence interval. As mentioned in point 2 above, I am not keen on the categorisation of gestational age and would either expect to see gestation modelled as a continuum (plus interactions) or full justification of why this was not done.

RESPONSE: We do not believe that the maturation of the foetal lung (or immune sytem) is linear by gestational age in relation to subsequent risk of airway disease and therefore have used the conventional gestational age categories. Also in terms of clinical interpretation, a continuous measure of gestational age would be less helpful, as it is only among the extremely premature infants where there is a notable interaction and these infants are at greater risk. We have added further clarification of the methods in the statistical analysis section, added the interaction term (and confidence interval) to the results section, as well as clarifying the interpretation of these findings in the results section.