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. Some articles will have been accepted based in part or entirely on reviews undertaken for other BMJ Group journals. These will be reproduced where possible.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Maternal and Perinatal Risk Factors for Childhood Cancer: Record Linkage Study
AUTHORS	Bhattacharya, Sohinee; Beasley, Marcus; Pang, Dong; Macfarlane, Gary

VERSION 1 - REVIEW

REVIEWER	Richard J.Q. McNally Reader in Epidemiology Institute of Health and Society Newcastle University UK
REVIEW RETURNED	14-Aug-2013

GENERAL COMMENTS	GENERAL This is an interesting and well written paper. However, there are a few important issues that require further comment. The conclusions regarding c-section are too strong. Childhood cancer has been associated with larger babies and this could be one reason for the apparent association. Furthermore, given that sub-type was not analyzed it is rather dangerous to infer too much from this tentative finding. SPECIFIC 1. The findings regarding Caesarian section should be treated with caution. The discussion and conclusions need to be greatly toned down. 2. The previous evidence regrading smoking is mixed. Reference 9 is not the correct one for the major UK Childhood Cancer Study. This needs to be corrected to give the citation for the full original manuscript (not subsequent correspondence relating to hepatoblastoma). It should be Br J Cancer 2003;88(3):373-381 3. Another reference that should be added and discussed is Dorak MT Cancer Causes Control 2007;18(2):219-228 Minor 1. Reference 19 also says "22"

REVIEWER	Charles Stiller Childhood Cancer Research Group University of Oxford New Richards Building Old Road Campus No competing interests
REVIEW RETURNED	19-Sep-2013

GENERAL COMMENTS	The authors have carried out a record-based case-control study of childhood cancer in relation to a range of possible maternal and perinatal risk factors. The principal finding is an approximately doubled odds ratio with delivery by caesarean. This association is unexplained but the authors suggest that epigenetic changes also associated with caesarean delivery should be investigated. Major comments 1. The study used high-quality record-based data and was appropriately analysed. The low number of cases was a severe limitation, making it impossible to analyse even sizeable and themselves heterogeneous subgroups (such as all leukaemias, all CNS tumours) within the extremely heterogeneous range of all childhood cancers. Even worse, Table 4 indicates that a considerable number of 'cases' were not in fact childhood cancer as defined, for example, in the standard International Classification of Childhood Cancer (ICCC) (Steliarova-Foucher et al., 2005), which does not include any of the non-malignant diagnoses listed under ICD-10 D40, D44, D47 and D48 (which between them account for 20 cases, 11% of the total of 176). This accounts in part for the staggeringly high cumulative incidence of 4.05 per 1000 up to age 15 years (Results, line 8), or 1 in 247; a cumulative rate of up to 1/400 (which is consistent with the quoted incidence rate for Scotland in 2007) is much more usual. The analyses should be repeated only for cases that have ICD-O codes mapping to ICCC categories. 2. To avoid the possibility of apparent risk factors actually being a direct or indirect result of a cancer already present at birth in some cases, the data should also be analysed excluding cases diagnosed within, say, the first few months of life. This could turn out to be particularly important for caesarean delivery. Do the authors have a detailed distribution of age at diagnosis for cases diagnosed in the first year? 3. Conclusion. The finding on caesarean section is not itself 'novel', only the size of the OR. Other comments 3. Introduction, lines 9-11. Is it really surprising that 'authors report conflicting results'? The more something is studied, especially if studies are small, the greater the opportunity for this to happen. 4. Methods - Data Sources, line 1-2. Please define more precisely. Are these births that took place in Aberdeen, or births where the mother was a resident of Aberdeen, or births where mother was an Aberdeen resident and gave birth there? 5. Methods - Data Linkage, penultimate sentence. Please give % with no CHI number. 6. Methods – Study Design, lines 3-5 from end. Down syndrome is also associated with AML/MDS (now known as ML of DS). 7. Methods – Study Design, final sentence. In fact it would only have been necessary to exclude children who died before the age at which the case to which they would potentially matched was diagnosed with cancer. This wider exclusion of any who died before age 15 could be a source of bias. Please justify. 8. Discussion. On Apgar score, please also discuss the recent paper by Li et al., BMJ Open 2012. 9. Discussion, page 10, middle of main paragraph. On mode of delivery, in addition to references 7, 35 and 36, please also comment on references 10 (lymphoma) and 3 (neuroblastoma). 10. Table 1. What does 'single/widowed' mean as a maternal or paternal social class?

VERSION 1 – AUTHOR RESPONSE

Reviewer: Richard J.Q. McNally
Reader in Epidemiology
Institute of Health and Society
Newcastle University
UK

GENERAL This is an interesting and well written paper. However, there are a few important issues that require further comment. The conclusions regarding c-section are too strong. Childhood cancer has been associated with larger babies and this could be one reason for the apparent association. Furthermore, given that sub-type was not analyzed it is rather dangerous to infer too much from this tentative finding.
We thank the reviewer for his positive comments and his observation that larger babies may be the link in the association between caesarean delivery and childhood cancer seen in this study. However, we adjusted for birthweight in the model and the association remained.
SPECIFIC
1. The findings regarding Caesarian section should be treated with caution. The discussion and conclusions need to be greatly toned down.
We agree with the reviewer and have done so in the revised draft.
2. The previous evidence regrading smoking is mixed. Reference 9 is not the correct one for the major UK Childhood Cancer Study. This needs to be corrected to give the citation for the full original manuscript (not subsequent correspondence relating to hepatoblastoma). It should be Br J Cancer 2003;88(3):373-381
Many thanks for correcting the reference
3. Another reference that should be added and discussed is Dorak MT Cancer Causes Control 2007;18(2):219-228
We thank the reviewer for his suggestion and have now added this reference to our discussion.

Minor
1. Reference 19 also says "22"
Many thanks for pointing out this error. This has now been corrected.

Reviewer: Charles Stiller
Childhood Cancer Research Group
University of Oxford
New Richards Building
Old Road Campus
No competing interests

The authors have carried out a record-based case-control study of childhood cancer in relation to a range of possible maternal and perinatal risk factors. The principal finding is an approximately doubled odds ratio with delivery by caesarean. This association is unexplained but the authors suggest that epigenetic changes also associated with caesarean delivery should be investigated.
Major comments
1. The study used high-quality record-based data and was appropriately analysed. The low number of cases was a severe limitation, making it impossible to analyse even sizeable and themselves heterogeneous subgroups (such as all leukaemias, all CNS tumours) within the extremely heterogeneous range of all childhood cancers. Even worse, Table 4 indicates that a considerable number of 'cases' were not in fact childhood cancer as defined, for example, in the standard International Classification of Childhood Cancer (ICCC) (Steliarova-Foucher et al., 2005), which does not include any of the non-malignant diagnoses listed under ICD-10 D40, D44, D47 and D48 (which between them account for 20 cases, 11% of the total of 176). This accounts in part for the

staggeringly high cumulative incidence of 4.05 per 1000 up to age 15 years (Results, line 8), or 1 in 247; a cumulative rate of up to 1/400 (which is consistent with the quoted incidence rate for Scotland in 2007) is much more usual. The analyses should be repeated only for cases that have ICD-O codes mapping to ICCC categories.
As suggested by the reviewer, we have now repeated the analyses excluding these ICD codes with minimal change in the results.
2. To avoid the possibility of apparent risk factors actually being a direct or indirect result of a cancer already present at birth in some cases, the data should also be analysed excluding cases diagnosed within, say, the first few months of life. This could turn out to be particularly important for caesarean delivery. Do the authors have a detailed distribution of age at diagnosis for cases diagnosed in the first year?
We had the date of birth and the date of cancer registration. On further scrutiny, all dates of cancer registration were after the first year of birth, except for one. We take the point of the reviewer, but felt that it would not be efficient to repeat analyses excluding the one case. Instead we have added a sentence in the discussion to take note of this.
3. Conclusion. The finding on caesarean section is not itself 'novel', only the size of the OR.
We have modified the conclusion according to the suggestions of the other reviewer.
Other comments
4. Introduction, lines 9-11. Is it really surprising that 'authors report conflicting results'? The more something is studied, especially if studies are small, the greater the opportunity for this to happen.
We agree with the reviewer and have removed the word "yet" from the sentence.
5. Methods - Data Sources, line 1-2. Please define more precisely. Are these births that took place in Aberdeen, or births where the mother was a resident of Aberdeen, or births where mother was an Aberdeen resident and gave birth there?
These are all births occurring in Aberdeen maternity Hospital, the only maternity hospital serving the population of Aberdeen city and district. We have added a sentence to clarify this.

6. Methods - Data Linkage, penultimate sentence. Please give % with no CHI number.
This has now been added.
7. Methods – Study Design, lines 3-5 from end. Down syndrome is also associated with AML/MDS (now known as ML of DS).
We thank the reviewer for pointing this out and have added this condition.
8. Methods – Study Design, final sentence. In fact it would only have been necessary to exclude children who died before the age at which the case to which they would potentially matched was diagnosed with cancer. This wider exclusion of any who died before age 15 could be a source of bias. Please justify.
We apologise for this confusion. As this was a case control study nested within a very large cohort, it was simpler to exclude all children who died of non-cancer causes at the outset. We think this is unlikely to lead to any selection bias in view of the very small proportion of potential controls that have been excluded.
9. Discussion. On Apgar score, please also discuss the recent paper by Li et al., BMJ Open 2012.
This has now been added.
Discussion, page 10, middle of main paragraph. On mode of delivery, in addition to references 7, 35 and 36, please also comment on references 10 (lymphoma) and 3 (neuroblastoma).
This has been added.
10. Table 1. What does 'single/widowed' mean as a maternal or paternal social class?
We apologise for this oversight. The AMND records social class as the Registrar general's occupation based social class based on the husband/ partner's occupation. Where the mother was single or widowed, the woman's own occupation was taken into consideration.

VERSION 2 – REVIEW

REVIEWER	Richard McNally Institute of Health & Society, Newcastle University, UK
REVIEW RETURNED	01-Nov-2013

GENERAL COMMENTS	The revision is confusing. There are a number of versions without consistency between them. Even though the authors state they have addressed my comments this has not been done. This paper requires very careful further revision. The revision is confusing. There are a number of versions without consistency between them. Even though the authors state they have addressed my comments this has not been done. This paper requires very careful further revision. It is not clear that the conclusions have been toned down. Perhaps a statement indicating that since it was not possible to analyze specific diagnostic groups caution should be exercised regarding the general nature of the findings (i.e. it not clear which specific diagnostic groups are behind the reported effects). Reference 9 is still not correct. It should be Pang D, McNally R, Birch JM. Parental smoking and childhood cancer: results from the United Kingdom Childhood Cancer Study. Br J Cancer 2003;88:373-381. Even though the authors state that they have discussed the paper by Dorak MT et al (Cancer Causes Control 2007;18(2):219-228) this does not appear to have been done. Reference 19/22 - amended in some versions but not all. Minor Rates should be reported as "per million children per year", not as "child-years" (e.g. Introduction, second sentence).

VERSION 2 – AUTHOR RESPONSE

Reviewer: 1

Comments to the Author

GENERAL: The authors present an interesting analysis of maternal and perinatal risk factors for childhood cancer. Unfortunately the present version of the paper suffers from a number of major shortcomings. There is a lack of clarity and some confusion in the arguments presented. For example, although the authors have found that the apparent effect of maternal smoking is not present after adjustment for maternal age, they persist in reporting this specific effect as a major "conclusion" from the study. Furthermore, they have not investigated if maternal age is significant after adjusting for maternal smoking. Also, later they explain that smoking status is absent for a large number of participants. Surely, this is likely to bias the results? There is a lack of explanation of the mechanisms that may play a role in aetiology. Statistical analyses should be driven by clearer hypotheses for testing. Results should follow logically from these analyses. At present they present a somewhat confused picture. The Discussion also needs a thorough revision. Specific issues are detailed below.

We thank the reviewer for their suggestions and have now presented the proportion of missing data for all variables and included maternal smoking in the multivariable model. We have also revised the discussion section.

SPECIFIC:

1. Abstract, page 2, Conclusions. This seems to be a re-statement of the results. How can the results be interpreted? What are the mechanisms? What are the next steps in the research?
   The conclusions have now been changed to take account of this.
2. Introduction, page 3, lines 25-27. The authors state that "...specific central nervous system tumours have been found to be associated with fetal growth..." What was the direction of the effect? This should be explained.
   This has now been clarified.
3. Introduction, page 3, lines 33-35. What are the "proposed epigenetic associations with cancer"? Give more details and add references.
   This has been clarified in the discussion and deleted from the introduction to avoid repetition.
4. Statistical Analysis, page 5, lines 43-47. What is meant by "small (less than 1/1.5) central values"? This is not clear.
   This has now been changed to include all variables.
5. Statistical Analysis, line 52. What is meant by "too small"? How small? This should be quantified.
   Please see response to comment no. 5.

6. Statistical Analysis, lines 52-54. What was the proportion of missing values for "smoking status" and also for "other variables". These data should be presented.

   These have now been presented in Table 1.
7. Results, page 60, lines 54-58 & page 7, lines 3-5. The authors state that "maternal smoking was not included as there was a large correlation between maternal age and smoking..." However, it is not clear if maternal age or smoking is underlying the associations. Further justification should be given. There is a general lack of clarity concerning these analyses.
   All variables have now been included in the final model and the mutually adjusted odds ratios presented in the table.

8. Discussion, page 7, lines 56-58. I am not clear why the authors persist in mentioning the smoking association as a conclusion, given that they state here that it is not significant after adjustment for maternal age.

This has now been modified to reflect the findings from the multivariable model.

9. Discussion, page 9, lines 13-19. Given that non-response bias is not applicable to the present study, what is the explanation? More explanation is needed.

10. Discussion, page 10, lines 35-37. The statement that the "association of caesarean delivery....is likely to have far reaching consequences for obstetric practice" is far too strong and should be removed or toned down considerably. First the finding would need to replicated in other larger studies. Then clear mechanisms would need to be identified.

We agree with the reviewer regarding replicating the study in larger population based cohorts, but maintain that the strong association found between caesarean delivery and childhood cancer are unlikely to be due to chance. We have discussed the possible mechanisms underlying the association and possible clinical implications.

11. Conclusion, page 11. Given my earlier comments, this needs to be re-written. There is a general lack of clarity and so the message is somewhat weak and obscure. What do these results mean? Is smoking related to increased risk? Or is it maternal age? What are the possible mechanisms?

This has now been changed.

12. Table 1. Give full details of missing numbers.

This has now been added.

Reviewer: 2

Comments to the Author

This study presents a seemingly straightforward analysis of linked registry data to investigate associations between childhood cancer and maternal and perinatal risk factors. However, there are some issues with the data collection, presentation and analysis that require attention before the conclusions can be made. Specific points to be addressed are outlined below:

1. Multivariate used when multivariable meant.

This has now been changed.

2. Deterministic matching was used to form links, but probabilistic linkage may yield a larger and less biased dataset. It seems that the deterministic matching was based solely on CHI number, please clarify. What is the sensitivity and specificity of CHI for matching in this group? Why wasn't probabilistic matching used? What evidence is there that the sample is not biased by the process used for linkage?

Deterministic linkage using CHI number where available has an accuracy of 97% to 99.8%. Where CHI number was unavailable, probabilistic matching was used. This has now been clarified in the text.

3. How were the 4 controls selected from the pool of available suitable matches for each case?

Wherever possible, 4 controls matched on sex were selected from the pool of non-cases who had their dates of birth closest to the date of birth of the case (2 before and 2 after).

8. Could the exclusion of children who died of non-cancer causes prior to their 15th birthday have caused any bias? The predictive model built will apply only to the subgroup who will not die from other causes. Assuming that the model is not meant to be used predictively (the outcome of death from other causes will not be known at birth/in early childhood), this may not be a problem, but it would be good to discuss any potential implications of this decision.
We felt that including children who could in no way have been at risk of developing cancer – stillbirth is an extreme example, would in fact bias our results from a retrospective analysis. Had we done a prospective (survival) analysis, these cases would have been censored.

5. The reliability of the information collected from the records should be discussed.

This has now been done.

9. It is not good practice to only consider variables that are univariately associated (or have a large or small OR) with outcome in the multivariable model. Some variables may only be associated after adjustment for others. All variables of interest should be considered for inclusion.
This has now been done.

7. It appears that maternal age is entered 'as a proxy for smoking' due to the high correlation. This seems quite an odd and somewhat unreasonable thing to do. The two variables may be correlated but that does not infer that one can negate the need for the other. It seems that the models were fitted and then the smoking variable removed because it had a larger number of missing values. How were missing values for any variables dealt with in the model building process? If complete case models were used then they would be based on varying numbers, making comparison between models difficult and potentially introducing bias. Some attempt should be made to replace missing data, perhaps via multiple imputation and the likely biases in the group with all variables recorded discussed.

Nowhere are we even told how many missing values there were nor what the correlation was between smoking and age.

This has now been included and explained in the text.

8. Maternal age is included in the final model yet the OR are not given in table 2. There does not seem to have been any investigation of whether age and smoking are independently significantly associated with outcome, nor for the other factors in table 1.
This has now been changed.

Reviewer: 3

Comments to the Author

This small record linkage study examines the relation between various routinely collected pregnancy/obstetric items and the subsequent development of childhood cancer. It is nicely presented and clearly written. Other than this I'm afraid I can say very little that is positive. It is an extremely small series; and I'm sorry to say that with only 176 cancers diagnosed 1993-2006 in their region in total, and consequently no examination by cancer subtype etc., it is underpowered and pointless. Incidentally, cases diagnosed 1991-94 will have been included in the national study that accrued cases during this period. Sorry to be so negative – but it adds nothing.

We are sorry that the reviewer feels that this small scale study adds nothing to the literature. Childhood cancer is a rare condition and it is therefore unlikely that the number of cases will be large. However, we felt that this small scale study had some interesting findings worth publishing and merit further investigation in a large scale population based cohort in the future.