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)	Associations between birth at, or after, 41 weeks gestation and perinatal encephalopathy: A Cohort Study.
AUTHORS	Odd, David; Yau, Christopher; Winter, Cathy; Draycott, Tim; Rasmussen, Finn

VERSION 1 - REVIEW

REVIEWER	Hutchon, David Emeritus Consultant Obstetrician, Memorial Hospital, Darlington, UK Competing interests: No competing interests
REVIEW RETURNED	28-Mar-2017

GENERAL COMMENTS	The size of the database is very impressive. the hypothesis of the study is that there will be differences in the parameters of the pregnancy outcomes after their due date. The results showed a statistically significant increase in encephalopathy after 41 weeks. They conclude "Expediting birth at 40 weeks gestation, including induction of labour, could prevent a substantial proportion (up to 5%) of all neonatal encephalopathy." however they have not included in their data whether any of these pregnancies were induced or indeed what the induction policy was over the 40 years that the data was collected. Were only spontaneous labours included ? They do not comment on the significantly lower caesarean section rate after 41 weeks. Although it is likely that most pregnancies were dated by early ultrasound in the later years, there is no mention about how the pregnancies were dated and it is likely that there were same changes over the 40 years. Was this data available form the database ? The statement "Whilst more research is done in this area, if circumstances are such that gestational age can be reliably measured, then membrane sweeping, which is not considered part of the formal induction process, could be introduced routinely from 38-39 weeks gestation to encourage spontaneous labour before 40 weeks gestation." is not acceptable. Membrane sweeping may not be considered part of formal induction but it is an intervention which is intended to result in the onset of labour. Such an intervention is only justified if it has been shown to reduce the risk of morbidity or death to the baby or mother.

REVIEWER	Shah, Vibhuti Mount Sinai Hospital, Toronto, Ontario Canada Competing interests: None
REVIEW RETURNED	24-Apr-2017

GENERAL COMMENTS	In this manuscript the authors report on the associations of being born at more than or equal to 41 weeks gestation (singleton infants) and perinatal encephalopathy using data from the nationwide birth registry of Sweden over a period of 39 years. The authors attempt to adjust for confounding factors. They conclude that infants born at or after 41 weeks gestation are at higher risk of having low Apgar scores and perinatal encephalopathy and that the association appeared to be more marked in older mothers and that these information can be used for decision making (i.e. induced labor and deliver the infants. This is an important work as the consequences of HIE are devastating both for the child and family and its impact on healthcare resource utilization. Major comments: 1) The study is conducted over a period of 39 years and the obstetrical management/guidelines have changed over time especially if a woman is post-term, i.e. at 42 weeks gestation. In table 2 the authors report on Apgar scores based on GA and it is very surprising to see babies being delivered at 43, 44 and 45 weeks. It would be interesting to see the denominators for each GA groups and if the authors can identify the reasons for being delivered at those GA is available it would be very interesting. In addition it would be interesting to see the trend in the GA at delivery over the almost 40 years period. 2) Abstract: a) Background section, lines 4-5L: The authors state that the aim of this work is to quantify the risks of infants' developing encephalopathy when birth occurs post-term and ..... the relationship. By definition post-term means infants delivered after 42 weeks gestation which does not align with the title. The authors also use this term in the manuscript at various places and does not align with the goal/aim. b) Methods: The dataset contains information on 4,036, 346 infants .... 1973-2012. This number not match with the results section of the baby nor with the title in Table 1(n-3,427,450). The authors should considering presenting a flow diagram and explaining the sample size, the number of infants excluded and reasons for exclusion and the final number used in the analysis. c) The primary outcomes is neonatal encephalopathy (seizures, encephalopathy or brain injury caused by asphyxia or with unspecified cause) is not clear. Seizures can occur for example without encephalopathy and did this definition vary over different time points, i.e. how were these diagnosis recorded over the last 40 years and how accurate is the information? d) Results section, last line: Just need to give p value no need to use the term interaction. Do the authors have results on long-term outcomes of these babies as they are very relevant for counseling-important for parents to know despite low Apgar score etc how many of them have had a normal outcome? 3) Manuscript Background (page 4): 1) Line 7, infants born after their due date (post-term)- infants born after their due date would be 40 weeks plus while post-term is after

	42 weeks.... 2) Line 15: The authors need to explicitly state how they define perinatal asphyxia as they correctly state that not all infants develop hypoxic-ischemic encephalopathy. 3) Lines 42-43: The authors state that the aim of this work is to quantify the risk of developing HIE or being born in poor condition.. What criteria were used to categorize an infant with HIE needs to be stated (e.g. the American College of Obstetrics and Gynecology and American Academy of Paediatrics have strict criteria. Also "" infant born in poor condition" is not scientific- please state explicitly what is mean (e.g. criteria to define the condition of the infant" Population (page 4): 1) Line 50: regard dataset It would be great if the authors can present a figure: Flow diagram of the patient population to say how many infants were included initially and how many excluded including reasons rather than describing in the text 2) Page 5, lines 16-18: The authors define HIE as either having seizures, encephalopathy or brain injury caused by asphyxia or with unspecified cause). As previously stated this would not be the ACOG/AAP definition of HIE. Seizures or encephalopathy can be due to various reasons including metabolic causes - how did the authors differentiate that from HIE? Were the long-term outcomes on these infants available in the dataset? 3) Line 42: Please correct the spelling of word "beyond" Results: 1) Page 6, lines 40-41 and Table 1 regarding maternal and neonatal infection: What definitions were used to diagnose infection? did they have to be blood culture positive? did the mothers have chorioamnionitis? as shown in table the risk of infection was higher in infants 41 weeks plus. The baseline characteristics are different and what was the impact of infection on the baby e.g. did these infant develop meningitis which could potentially lead to seizures. 2) Table: Outcomes measures split by GA It appears from the table that some infants were delivered at 43, 44 and 45 weeks which is surprising? How many infants were delivered at GA/ Would be interesting to see in a subgroup analysis of these infants how many had HIE? Discussion (page 8): As the authors point that infants born at extreme of GA < 37 weeks or more than or equal to 41 weeks had higher risk of low Apgar scores and/or HIE. Line 46 they suggest that the reason for poor outcome may be due deterioration in utero environment but do not elaborate on what factors may lead? could this be due to placental insufficiency. Did the authors have access to placental pathology for these infants so that the data can be compared for infants < 41 weeks to > 41 weeks. 5) Based on these findings what would the authors recommend in regards to change in clinical practice?

REVIEWER	Koutoumanou, Eirini UCL GOS Institute of Child Health London, UK Competing interests: none
REVIEW RETURNED	16-Jun-2017

GENERAL COMMENTS	This is a well written and interesting report and could improve with few (not major) corrections/edits: - The authors mention that “infants born ≥41 week gestation were more likely to be female…” – but Table 1 shows that 52.6% were male, so shouldn’t it be the other way round? Or is there a mistake/misunderstanding somewhere? - The authors have extrapolated a 3% risk increase of encephalopathy per day overdue from the 1.22 OR comparing extra weeks (22% per extra week), but I am not confident that this extrapolation is valid and I would suggest it is removed and the authors cite the weekly increase only. - I would recommend that the authors add some comments as to why most of the results of the 45 week gestation group seem to be the same or better compared to the younger than 40/41 weeks group. Risk of encephalopathy for example was reduced after week 42. - The authors mention in a couple of places that they found ‘…a disproportionately higher risk in women over 35 years old’. Is this referring to the ORs shown on lines 43-57 on page 7 of 18? <35 OR=1.34 vs >35 OR=1.67? Is it this difference in the ORs that the authors are deeming as disproportionate? - Some of the differences observed between the groups are only minor, therefore I would recommend that the authors stress that on their comments. For example, maternal age is different only by 0.3 and neonatal infection by 0.2%. - Finally, the authors close the discussion section with a remark regarding membrane sweeping, which I believe is not overall appropriate. It has not been mentioned at any earlier part of the report that this particular technique was of interest and it raises more questions than it provides answers, hence I would recommend that this comment is removed/moved to the final remark of the conclusion section as one of the techniques to be further investigated.

VERSION 1 – AUTHOR RESPONSE

Comments with Regards to the Review:

Reviewer: 1

Comments to the Author

The size of the database is very impressive. the hypothesis of the study is that there will be differences in the parameters of the pregnancy outcomes after their due date. The results showed a statistically significant increase in encephalopathy after 41 weeks. They conclude "Expediting birth at 40 weeks gestation, including induction of labour, could prevent a substantial proportion (up to 5%) of all neonatal encephalopathy." however they have not included in their data whether any of these pregnancies were induced or indeed what the induction policy was over the 40 years that the data was collected. Were only spontaneous labours included ? They do not comment on the significantly lower caesarean section rate after 41 weeks. Although it is likely that most pregnancies were dated by early ultrasound in the later years, there is no mention about how the pregnancies were dated and it is

likely that there were same changes over the 40 years. Was this data available form the database ?
> We unfortunately have limited data on the induction of labour, or the reasons for caesarian section. Equally the gestational age was derived from the clinical notes, and the methods to derive this would have changed over the study period. We have added some clarification to the paper regarding these points.

The statement "Whilst more research is done in this area, if circumstances are such that gestational age can be reliably measured, then membrane sweeping, which is not considered part of the formal induction process, could be introduced routinely from 38-39 weeks gestation to encourage spontaneous labour before 40 weeks gestation." is not acceptable. Membrane sweeping may not be considered part of formal induction but it is an intervention which is intended to result in the onset of labour. Such an intervention is only justified if it has been shown to reduce the risk of morbidity or death to the baby or mother.
> We have modified the statement to a more measured conclusion with regard to this (and other) comments.

Reviewer: 2

Comments to the Author

In this manuscript the authors report on the associations of being born at more than or equal to 41 weeks gestation (singleton infants) and perinatal encephalopathy using data from the nationwide birth registry of Sweden over a period of 39 years. The authors attempt to adjust for confounding factors. They conclude that infants born at or after 41 weeks gestation are at higher risk of having low Apgar scores and perinatal encephalopathy and that the association appeared to be more marked in older mothers and that these information can be used for decision making (i.e. induced labor and deliver the infants. This is an important work as the consequences of HIE are devastating both for the child and family and its impact on healthcare resource utilization.
Major comments:
1) The study is conducted over a period of 39 years and the obstetrical management/guidelines have changed over time especially if a woman is post-term, i.e. at 42 weeks gestation. In table 2 the authors report on Apgar scores based on GA and it is very surprising to see babies being delivered at 43, 44 and 45 weeks. It would be interesting to see the denominators for each GA groups and if the authors can identify the reasons for being delivered at those GA is available it would be very interesting. In addition it would be interesting to see the trend in the GA at delivery over the almost 40 years period.
> We have added the absolute numbers to Table 2 as suggested (along with the proportion with outcome of low Apgar scores or encephalopathy). Medians', ranges, and IQR were similar for all years; so geometric means (with CI) have been presented to aid the reader' interpretation of temporal changes.

2) Abstract:
a) Background section, lines 4-5L: The authors state that the aim of this work is to quantify the risks of infants' developing encephalopathy when birth occurs post-term and ..... the relationship. By definition post-term means infants delivered after 42 weeks gestation which does not align with the title. The authors also use this term in the manuscript at various places and does not align with the goal/aim.
> We agree that the use of "post-term" does not align with the analysis and testing we have done here and have removed the references to clarify this.

b) Methods: The dataset contains information on 4,036, 346 infants .... 1973-2012. This number not match with the results section of the baby nor with the title in Table 1(n-3,427,450). The authors

should considering presenting a flow diagram and explaining the sample size, the number of infants excluded and reasons for exclusion and the final number used in the analysis.
> We have added an appendix with the dataset flow explained graphically as suggested.

c) The primary outcomes is neonatal encephalopathy (seizures, encephalopathy or brain injury caused by asphyxia or with unspecified cause) is not clear. Seizures can occur for example without encephalopathy and did this definition vary over different time points, i.e. how were these diagnosis recorded over the last 40 years and how accurate is the information?
> We have expanded the limitations section of the discussion to cover this and other points. We have changed to prose to reflect the primary outcome of likely perinatal encephalopathy (as referenced in the title). In particular we have expanded the prose to recognize the limitations of using a routine data source, and the possibility of misclassification: and the steps we have taken to minimise it.

d) Results section, last line: Just need to give p value no need to use the term interaction. Do the authors have results on long-term outcomes of these babies as they are very relevant for counseling- important for parents to know despite low Apgar score etc how many of them have had a normal outcome?
> We have amended the p-value as suggested.
> We do not have the ability to link this dataset with long term outcomes for the infants, but a complementary piece of work is ongoing with this in mind.

3) Manuscript
Background (page 4):
1) Line 7, infants born after their due date (post-term)- infants born after their due date would be 40 weeks plus while post-term is after 42 weeks....
> We agree that the use of “post-term” does not align with the analysis and testing we have done here and have removed the references to clarify this.

2) Line 15: The authors need to explicitly state how they define perinatal asphyxia as they correctly state that not all infants develop hypoxic-ischemic encephalopathy.
> We have expanded the limitations section of the discussion to cover the diagnostic criteria we have used for encephalopathy and have changed to prose to reflect the primary outcome of likely perinatal encephalopathy (as referenced in the title) rather than make the additional assumption of a perinatal/intrapartum asphyxia event.

3) Lines 42-43: The authors state that the aim of this work is to quantify the risk of developing HIE or being born in poor condition. What criteria were used to categorize an infant with HIE needs to be stated (e.g. the American College of Obstetrics and Gynecology and American Academy of Paediatrics have strict criteria. Also "infant born in poor condition" is not scientific- please state explicitly what is mean (e.g. criteria to define the condition of the infant)
> Please see the comment above. Much of the criteria used in this work has been chosen to be consistent with other work we have performed. The definition and process of defining encephalopathy has been expanded on, and the diagnostic limitation in this work expanded upon in the discussion. The phrase ‘poor condition’ has previously been used to capture this group of babies with low Apgar score, or other signs of compromise, without being over specific as to the underlying cause (so commonly assumed to be asphyxia). As we only use the Apgar scores in this work we have changed the wording to clarify this.

Population (page 4):
1) Line 50: regard dataset It would be great if the authors can present a figure: Flow diagram of the patient population to say how many infants were included initially and how many excluded including

reasons rather than describing in the text
> We have added an appendix with the dataset flow explained graphically as suggested. We have left the text in at present, but would be happy to remove it if the journal felt appropriate.

2) Page 5, lines 16-18: The authors define HIE as either having seizures, encephalopathy or brain injury caused by asphyxia or with unspecified cause). As previously stated this would not be the ACOG/AAP definition of HIE. Seizures or encephalopathy can be due to various reasons including metabolic causes - how did the authors differentiate that from HIE? Were the long-term outcomes on these infants available in the dataset?
> We have expanded the limitations section of the discussion to cover this and other points. In particular we recognize the limitations of using a routine data source, and the possibility of misclassification: and the steps we have taken to minimise it. We are unable to link this exact dataset with longer term outcomes, but further work looking at long-term outcomes is ongoing.

3) Line 42: Please correct the spelling of word "beyond"
> Corrected

Results:
1) Page 6, lines 40-41 and Table 1 regarding maternal and neonatal infection: What definitions were used to diagnose infection? did they have to be blood culture positive? did the mothers have chorioamnionitis? as shown in table the risk of infection was higher in infants 41 weeks plus. The baseline characteristics are different and what was the impact of infection on the baby e.g. did these infant develop meningitis which could potentially lead to seizures.
> The diagnosis in this work were derived from clinical codes (ICD 8-10) placed in the birth registry and as such the details of the clinical signs that led to them is not recorded. We have clarified this in the methods section. We have expanded the discussion about the increase risk of infection, as well as our adjustment for it in the final analysis.

Table: Outcomes measures split by GA
It appears from the table that some infants were delivered at 43, 44 and 45 weeks which is surprising? How many infants were delivered at GA/ Would be interesting to see in a subgroup analysis of these infants how many had HIE?
> We have added the absolute numbers to Table 2 as suggested (along with the proportion with outcome of low Apgar scores or encephalopathy).

Discussion (page 8):
As the authors point that infants born at extreme of GA < 37 weeks or more than or equal to 41 weeks had higher risk of low Apgar scores and/or HIE. Line 46 they suggest that the reason for poor outcome may be due deterioration in utero environment but do not elaborate on what factors may lead? could this be due to placental insufficiency. Did the authors have access to placental pathology for these infants so that the data can be compared for infants < 41 weeks to > We did not have access to any clinical data beyond that recoded in the birth registry, and have clarified this further in the text. We have also expanded the discussion around this point as suggested.

5) Based on these findings what would the authors recommend in regards to change in clinical practice?
> We have modified our clinical conclusions with regard to this and other comments.

Reviewer: 3

Comments to the Author

This is a well written and interesting report and could improve with few (not major) corrections/edits:

- The authors mention that “infants born ≥41 week gestation were more likely to be female…” – but Table 1 shows that 52.6% were male, so shouldn’t it be the other way round? Or is there a mistake/misunderstanding somewhere?
> This is an error in the interpretation and we have amended it in the paper.

- The authors have extrapolated a 3% risk increase of encephalopathy per day overdue from the 1.22 OR comparing extra weeks (22% per extra week), but I am not confident that this extrapolation is valid and I would suggest it is removed and the authors cite the weekly increase only.
> We have amended as suggested.

- I would recommend that the authors add some comments as to why most of the results of the 45 week gestation group seem to be the same or better compared to the younger than 40/41 weeks group. Risk of encephalopathy for example was reduced after week 42.
> We have amended as suggested.

- The authors mention in a couple of places that they found ‘…a disproportionately higher risk in women over 35 years old’. Is this referring to the ORs shown on lines 43-57 on page 7 of 18? <35 OR=1.34 vs >35 OR=1.67? Is it this difference in the ORs that the authors are deeming as disproportionate?
> The analysis suggested a modification of effect by maternal age (the two OR the referees has rightly identified). We have removed the word “disproportionate” from the work to make this clearer.

- Some of the differences observed between the groups are only minor, therefore I would recommend that the authors stress that on their comments. For example, maternal age is different only by 0.3 and neonatal infection by 0.2%.
> We have amended the discussion as suggested.

- Finally, the authors close the discussion section with a remark regarding membrane sweeping, which I believe is not overall appropriate. It has not been mentioned at any earlier part of the report that this particular technique was of interest and it raises more questions than it provides answers, hence I would recommend that this comment is removed/moved to the final remark of the conclusion section as one of the techniques to be further investigated.
> We have modified the statement to a more measured conclusion with regard to this (and other) comments.

VERSION 2 – REVIEW

REVIEWER	Shah, Vibhuti Mount Sinai Hospital, Toronto, Ontario Canada Competing interests: None
REVIEW RETURNED	27-Aug-2017

GENERAL COMMENTS	In this cohort study the authors use data on a cohort of infants born between 1973-2012 with the aim to quantify the risk of developing encephalopathy or being born with low Apgar scores when born 7 or more days after their due date. The authors report that the relative risk of developing encephalopathy increases by an estimated 20% per week after the due date and modified by maternal age. They

	conclude that these data can be used to provide counseling women as part of their decision making. Major comments: 1) Introduction section: The authors do not report on what the incidence of perinatal asphyxia is and its implications and also what proportion of infants truly end up with HIE as this is the group of babies with adverse neurodevelopmental outcome. 2) Lines 31-32: The authors state that the management of pregnancies beyond term varies hugely between units and countries. So what is know based on the studies from these countries. The authors make a statement without supporting references. 3) The outcome chosen by the authors is risk of perinatal encephalopathy or being born with low Apgar scores. Two points: a) the authors using ICD codes are including cases of perinatal encephalopathy of diverse etiology e.g. if the etiology is a metabolic cause then early delivery will have no impact on the outcome? So what information in regards to counseling the couple would one provide? b) low Apgar score- is this at 5 mins? 10 mins? as we know resuscitation and active management will alter the clinical course. If the infant does not improve despite resuscitation then the infant is most likely to develop HIE and its consequences. Methods: 1) on page 4 line 57: The authors used the data coded the diagnoses using ICD 8th, 9th and 10th revisions. They state that the details of clinical signs that led to the diagnosis are not recorded. Even if the clinical signs are not recorded the clinical diagnosis should be recorded. What would be valuable is the etiology of the encephalopathy and understanding which etiologies/factors can be modifiable? How does this inform counseling? 2) Page 5, lines 16: The primary outcome is the development of "likely perinatal encephalopathy" what is meant by likely? did these infants need admission to the NICU - if so would suggest that they were ill and have data on the etiology?? My major struggles continues to be the heterogeneous outcome definition used. Seizures may be due to HIE or infarction or stroke (not modifiable factor) as most have occurred prior to birth... 3) Section on potential confounders and covariates: Please support why these factors were chosen based on the literature and reference them appropriately. It maternal age is a confounder how will that be modified? same for maternal education? Results: 1) Page 6, line 58 to next page These infants were more likely to develop maternal or neonatal infection. No rationale has been provided in the discussion why this is so? Discussion: 1) It appears that the authors are recommending that expediting birth at 40 weeks including IOL could prevent a substantial proportion of encephalopathy. Based on the findings and conclusion of the study what changes will be made in the guidelines or in clinical practice??

REVIEWER	Koutoumanou, Eirini UCL, UK Competing interests: none

REVIEW RETURNED	05-Sep-2017

GENERAL COMMENTS	The last sentence of the abstract needs rephrasing: “These data could usefully be provided for women as part of their decision making.” to something such as “These data could be useful if provided to women…” or “could be useful for women…”. What does “ICD coding” stand for? “In total, 973,430 (28.4%) infants were born one or more weeks after their due date. Over the 40 years period, there was a slight reduction in the (geometric) mean of gestational age at birth from 40.1 (40.0-40.1) in 1973 to 39.6 (39.6-39.6) (p<0.001).” The 39.6 estimate is for when, i.e. 1973 vs what year? I believe a bit more explanation is needed for the fact that the Apgar 1minute score shows improvement after week 43 – is it not as accurate as the measurements at 5 and 10 minutes? Please correct “…on only 1 infants developing…” to “…only 1 infant...” Please change the labelling of Table 3 to be clear that the ORs are shown on the bottom bit of the table, whereas the rest of the values are mean differences.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 1

Comments to the Author
In this cohort study the authors use data on a cohort of infants born between 1973-2012 with the aim to quantify the risk of developing encephalopathy or being born with low Apgar scores when born 7 or more days after their due date. The authors report that the relative risk of developing encephalopathy increases by an estimated 20% per week after the due date and modified by maternal age. They conclude that these data can be used to provide counselling women as part of their decision making.

Major comments:
1) Introduction section:
The authors do not report on what the incidence of perinatal asphyxia is and its implications and also what proportion of infants truly end up with HIE as this is the group of babies with adverse neurodevelopmental outcome.
- We have added frequencies and references as suggested.

2) Lines 31-32: The authors state that the management of pregnancies beyond term varies hugely between units and countries. So what is know based on the studies from these countries. The authors make a statement without supporting references.
- We have added frequencies and references as suggested.

3) The outcome chosen by the authors is risk of perinatal encephalopathy or being born with low Apgar scores. Two points: a) the authors using ICD codes are including cases of perinatal encephalopathy of diverse etiology e.g. if the etiology is a metabolic cause then early delivery will have no impact on the outcome? So what information in regards to counseling the couple would one provide?

-The use of the routine data source and ICD codes does of course limit the interpretation of the study. We have further expanded the limitations section of the discussion and conclusion to reflect this.

b) low Apgar score- is this at 5 mins? 10 mins? as we know resuscitation and active management will alter the clinical course. If the infant does not improve despite resuscitation then the infant is most likely to develop HIE and its consequences.
- We have used a level of less than 7 at 5 minutes as our definition of a low Apgar score. We have defined this in the methods section and added the definition to the introduction to clarify this.

Methods:
1) on page 4 line 57: The authors used the data coded the diagnoses using ICD 8th, 9th and 10th revisions. They state that the details of clinical signs that led to the diagnosis are not recorded. Even if the clinical signs are not recorded the clinical diagnosis should be recorded. What would be valuable is the etiology of the encephalopathy and understanding which etiologies/factors can be modifiable? How does this inform counseling?
- As mentioned above the use of ICD codes does limit any further interrogation of the dataset, and we have added a paragraph to discussion this in the discussion.

2) Page 5, lines 16: The primary outcome is the development of "likely perinatal encephalopathy" what is meant by likely?
- We have amended the discussion to include and explain that some uncertainty (as in any study) exists around the diagnostic criteria used.

..did these infants need admission to the NICU - if so would suggest that they were ill and have data on the etiology??
- We don't have details on the place of care for these infants. We have added a comment reflecting this to the discussion, along with expansion on the discussion on our outcome definition. We have added further prose discussing the limitations of the ICD codes.

My major struggles continues to be the heterogeneous outcome definition used. Seizures may be due to HIE or infarction or stroke (not modifiable factor) as most have occurred prior to birth...
- This work does indeed have a number of limitations due to its use of routine collected data. However I hope we have balanced the statistical findings with an appropriate discussion for the reader to understand the interpretation, and limits of the work.

3) Section on potential confounders and covariates:
Please support why these factors were chosen based on the literature and reference them appropriately. It maternal age is a confounder how will that be modified? same for maternal education?
- We have defined the confounders a-prior based on presumed causal pathways, and have referenced a paper discussing the role and place of this in the adjusted analyses. The role of the adjusted analyses is to identify the possibility of an underlying causal pathway outside of possible confounders, rather than identify intervention points; although it should be noted that little evidence of confounding (e.g. deviation of the point estimates) was seen in this work.

Results:
1) Page 6, line 58 to next page These infants were more likely to develop maternal or neonatal infection. No rationale has been provided in the discussion why this is so?
- This is perhaps outside the analysis of the paper, although we have added some discussion of it to the paper, and the covariate was incorporated into the analysis to help control for this.

Discussion:

1) It appears that the authors are recommending that expediting birth at 40 weeks including IOL could prevent a substantial proportion of encephalopathy. Based on the findings and conclusion of the study what changes will be made in the guidelines or in clinical practice??
- We suggest that this data could be used when discussing the options for delivery in specific cases; and have expanded the conclusions with this in mind.

Reviewer: 2

Comments to the Author
Thank you for all the corrections and edits. The manuscript is improved. I have listed few minor comments below:

The last sentence of the abstract needs rephrasing: "These data could usefully be provided for women as part of their decision making." to something such as "These data could be useful if provided to women…" or "could be useful for women…".
- The prose has been amended as suggested

What does "ICD coding" stand for?
- International Classification of Disease: we have added this to the text.

"In total, 973,430 (28.4%) infants were born one or more weeks after their due date. Over the 40 years period, there was a slight reduction in the (geometric) mean of gestational age at birth from 40.1 (40.0-40.1) in 1973 to 39.6 (39.6-39.6) (p<0.001)." The 39.6 estimate is for when, i.e. 1973 vs what year?
- This was the last year (2012) of the data: we have amended the text appropriately

I believe a bit more explanation is needed for the fact that the Apgar 1minute score shows improvement after week 43 – is it not as accurate as the measurements at 5 and 10 minutes?
- It's unclear why this should happen, although statistical sampling or inaccuracies in the score are possible. We have added a sentence to discuss this in the discussion.

Please correct "…on only 1 infants developing…" to "…only 1 infant..."
- The prose has been amended as suggested

Please change the labelling of Table 3 to be clear that the ORs are shown on the bottom bit of the table, whereas the rest of the values are mean differences.
- The prose has been amended as suggested

VERSION 3 – REVIEW

REVIEWER	Koutoumanou, Eirini UCL, UK Competing interests: None
REVIEW RETURNED	26-Oct-2017

GENERAL COMMENTS	I have a list of few more corrections below, some of which, surprisingly, were actually part of my review for the very first version of this manuscript, so I am not sure how things got lost in the process. Specifically on this last point, the authors still mention the 3% increase in the risk of encephalopathy per day at the "What this

	study adds" section; which is the wrong interpretation of the OR of 1.22. Throughout the paper, I recommend that the word risk is replaced by odds when interpreting the results. An OR of 1.22 means higher odds, which also means higher risk but not by 22%. One would need to calculate the relative risk to find the direct increase in the risk. I still think that the authors have not put enough emphasis on the magnitude of some of the differences when interpreting. • Older mothers by 0.11 of a year are not really that much older. • All percentages shown on Table 1 regarding education status are very similar between the comparison groups. • Apgar scores at 5 and 10 minute differ between 0.02 to 0.06 units – is that a noteworthy change? The authors have added a comment about this at the results section, but I don't think it's a very successful one as it's followed by another comment which I don't think ties in nicely: "Of note however, while there were differences in demographics between the two groups investigated here, many were often of minor magnitude (e.g. a 2-3 month difference in maternal age). It seems possible therefore that induction of labour or caesarean section at 40 weeks gestation may improve morbidity and mortality, although we were unable to investigate perinatal deaths in this work." I am really not sure as to how the comment about magnitude relates to the perinatal deaths. Regarding Appendix 2, the authors should rephrase the way they refer in the text to the comparisons of the mothers' employment and education status. The comparison is made between those with and without missing data, e.g. per level of education and not amongst all levels of education. In infants with missing data, there were more mothers of all education levels, apart from full secondary, but the authors have interpreted this in the text as "…had less education…". Similarly for occupation, more mothers were both on manual and other. Please revisit. Could the authors please clarify further sensitivity analysis 5? Missing data on which variables were imputed? If one adds up the numbers shown on the column of "Infants with missing data" in Appendix 2, the result is about 600-700,000 infants. We are only told of infants with missing Apgar or gestational age or with missing at least one covariate were excluded. So which values have been imputed? I am also unclear as to what the following means: "Because of technical limitations only a random 10% of infants without encephalopathy were included in this analysis and the model weighted to represent the initial population." Could the authors please replace the p-values in Table 1 by confidence intervals, as due to the very large sample sizes, the p-values are bound to be significant? Can the authors please include the risk of encephalopathy between those with and without missing data in Appendix 2, instead of just reporting the p- value of the comparison in the text? "There was no difference in gender (p=0.726) or risk of encephalopathy (p=0.136) between those with or without complete data." Could the authors please clarify where the following statement is

	derived from? "..could prevent a substantial proportion (up to 5%) of all neonatal encephalopathy." Minor corrections: In Appendix 2, could the authors please double check the numbers of the maternal age between the two groups, 28 vs 28.2 but difference reported as 0.11? Please replace "were assumed" with "were treated as" at the following sentence: "Birthweight and maternal age were assumed to be continuous variable..." Please rephrase "While these results should perhaps not be viewed in isolation" to "While these results should not be viewed in isolation" – no results should ever be viewed in isolation

VERSION 3 – AUTHOR RESPONSE

I have a list of few more corrections below, some of which, surprisingly, were actually part of my review for the very first version of this manuscript, so I am not sure how things got lost in the process.

Specifically on this last point, the authors still mention the 3% increase in the risk of encephalopathy per day at the "What this study adds" section; which is the wrong interpretation of the OR of 1.22. Throughout the paper, I recommend that the word risk is replaced by odds when interpreting the results. An OR of 1.22 means higher odds, which also means higher risk but not by 22%. One would need to calculate the relative risk to find the direct increase in the risk.

- We agree with the reviewer that the odds is different from the risk: although in this work (with such a rare outcome) the two should approximate. We have however changed the prose as suggested where we referred to odds (rather than risk) as appropriate. However where we state 'higher' or 'lower' risk (and evidence by raised OR) we have retained 'risk' as this remains a valid statement and, we believe, is more intuitive to the reader. Changing to 'odds' would also be correct and we would be happy to do so if the editors felt this improved the text. We have amended the statement in the "What this study adds" section to remove the 3% per day text.

I still think that the authors have not put enough emphasis on the magnitude of some of the differences when interpreting.

- Older mothers by 0.11 of a year are not really that much older.
- All percentages shown on Table 1 regarding education status are very similar between the comparison groups.
- Apgar scores at 5 and 10 minute differ between 0.02 to 0.06 units – is that a noteworthy change?

The authors have added a comment about this at the results section, but I don't think it's a very successful one as it's followed by another comment which I don't think ties in nicely: "Of note however, while there were differences in demographics between the two groups investigated here, many were often of minor magnitude (e.g. a 2-3 month difference in maternal age). It seems possible therefore that induction of labour or caesarean section at 40 weeks gestation may improve morbidity and mortality, although we were unable to investigate perinatal deaths in this work." I am really not

sure as to how the comment about magnitude relates to the perinatal deaths.

- We agree that the mean difference of Apgar scores may not be an important difference in itself, although the proportion with a low score is perhaps more important and we have changed the statements and drawn attention in the results to the small magnitude of differences seen. We have also (as suggested below) amended table 1 to present absolute differences in measures (rather than p-values), helping the reader can better interpret if the difference is important.

We have also modified the paragraph regarding the small differences in clinical measures between exposed and unexposed infants and moved it to the paragraph discussing confounders, rather than placing it within the section discussing the outcomes.

Regarding Appendix 2, the authors should rephrase the way they refer in the text to the comparisons of the mothers' employment and education status. The comparison is made between those with and without missing data, e.g. per level of education and not amongst all levels of education. In infants with missing data, there were more mothers of all education levels, apart from full secondary, but the authors have interpreted this in the text as "…had less education…". Similarly for occupation, more mothers were both on manual and other. Please revisit.

- We have changed the text to reflect the complex changes in occupation etc. seen in the missing data, and clarified the format of the table (and which infants were in which cohort).

Could the authors please clarify further sensitivity analysis 5? Missing data on which variables were imputed? If one adds up the numbers shown on the column of "Infants with missing data" in Appendix 2, the result is about 600-700,000 infants. We are only told of infants with missing Apgar or gestational age or with missing at least one covariate were excluded. So which values have been imputed? I am also unclear as to what the following means: "Because of technical limitations only a random 10% of infants without encephalopathy were included in this analysis and the model weighted to represent the initial population."

- All variables, except exposure and outcome variables were imputed, and we have now clarified this in the prose. Missing data was clustered and so many infants had more than one missing data point. Appendix 2 aims to present the characteristics of those infants included or excluded from the main analyses. We have added a denominator at the top of the table to clarify this. We used a random 10% sample as a fully imputed dataset provided substantial computational issues (having over 71 million subjects with multiple data points). We have added clarification of this to the text.

Could the authors please replace the p-values in Table 1 by confidence intervals, as due to the very large sample sizes, the p-values are bound to be significant?

- Amended as suggested

Can the authors please include the risk of encephalopathy between those with and without missing data in Appendix 2, instead of just reporting the p-value of the comparison in the text? "There was no difference in gender (p=0.726) or risk of encephalopathy (p=0.136) between those with or without complete data.

- Amended as suggested

Could the authors please clarify where the following statement is derived from? “..could prevent a substantial proportion (up to 5%) of all neonatal encephalopathy.”

- It’s based on the calculated estimated population attributable risk fraction, and we have added clarification of this to the prose.

Minor corrections:

In Appendix 2, could the authors please double check the numbers of the maternal age between the two groups, 28 vs 28.2 but difference reported as 0.11?

- We have checked and the numbers are correct. The detailed results are;

Study Cohort: 28.04315 (28.03768 to 28.04862)

Missing: 28.15793 (28.13766 to 28.17819)

Difference: 0.1147773 (0.0950165 to 0.134538)

Please replace “were assumed” with “were treated as” at the following sentence: “Birthweight and maternal age were assumed to be continuous variable...”

- Amended as suggested

Please rephrase “While these results should perhaps not be viewed in isolation” to “While these results should not be viewed in isolation” – no results should ever be viewed in isolation

- Amended as suggested