Author's response to reviews

Title: Does induction of labor for constitutionally large-for-gestational-age fetuses identified in utero reduce maternal morbidity? A historical cohort study.

Authors:

Françoise Vendittelli (fvendittelli@chu-clermontferrand.fr)
Olivier Rivière (olivier.riviere@audipog.net)
Brigitte Neveu (Brigitte.neveu@imm.fr)
Didier Lémery (dlemery@chu-clermontferrand.fr)

Version: 2
Date: 8 March 2014

Author's response to reviews: see over
To: The Editor in Chief of BMC pregnancy and childbirth

Clermont-Ferrand, March 6, 2014

Dear Sir,

We thank the reviewers for their thorough comments, which were globally favorable to our study, "Does induction of labor for constitutionally large-for-gestational-age fetuses identified in utero reduce maternal morbidity? A historical cohort study by Françoise Vendittelli, Olivier Rivière, Brigitte Neveu and Didier Lémery (MS: 1017138821111671).

Modifications to the article are in red in the article.

Reviewer #1: Clare Tower

Reviewer's report:
The issue of how best to manage large babies identified on ultrasound scan in terms of planning delivery is a question that remains unanswered. This study is a retrospective database analysis comparing induction vs continuing pregnancy when a large baby is identified on scan at 37-38 weeks. As such, I think it makes a valuable contribution to the literature in this area. However, as the paper stands it needs some considerable improvement before publication, largely around making the presentation clearer, more concise and more discussion around the bias inherent in it.

Major Compulsory revisions
• The last sentence of the background section of the abstract does not make sense. It needs to state: ‘This study aimed to ……’

Answer to reviewer: We have made this correction in the text.

• The first sentence of the methods section does not make sense. In addition, the methods do not really summarise how the study was done – needs to state retrospective study on French database etc.

Answer to reviewer: The methods section clearly states that this is a retrospective study -- which is a part of what the term historical study means. It is clearly specified that we are using a perinatal database. We have modified the formulation, using shorter and we hope clearer sentences.
• Conclusion in abstract not correct – this states that identification in utero of large babies does not reduce maternal morbidity. I thought the study was comparing induction with not at 37-38+6 weeks, not identification

Answer to reviewer: We agree with the reviewer and have made the correction.

Introduction
• The introduction needs to discuss and define what exactly is meant by severe perineal lacerations – this would usually by 3/4th degree tears but there is no discussion of these tears at all in the manuscript (although they are in the table). Most obstetricians/gynaecologists would consider that this is where most of the morbidity lies.

Answer to reviewer: We have added the definition of severe perineal lacerations to the text, as requested.

The following sentence needs explanation: ‘Studies have focused on the utility of inducing labor retrospectively, that is, based on known birth weight [16,17].’ How can labour be induced retrospectively?

Answer to reviewer: The sentence has been rewritten. “Studies examining the utility of inducing labor have generally used a retrospective analysis based on actual birth weight, rather than estimated in utero fetal weight [16,17].”

The English is poor in this sentence: ‘Some studies have examined the interest of inducing labor for fetuses with macrosomia suspected in utero [18-23].’ – What is meant by ‘the interest of’ – do you mean benefit of?

Answer to reviewer: Yes, we have modified the sentence.

Methods
Please re-write the 1st sentence of the methods section entitled ‘definitions of variables and statistical analysis’. I was confused by this as it makes it sound as though the comparison is between induction of labour or spontaneous labour at the gestation given. Actually the comparison is induction of labour vs expectant management / continuing pregnancies at these gestations.

Answer to reviewer: We have corrected the sentence, as requested.

Results
Did the authors check the data was normally distributed before using means to describe the data?

Answer to reviewer: The hypothesis of normality for a test comparing two means is secondary once the number of individuals has reached at least 30, which is the case here (see statistical textbooks, for example, Bouyer J et al. Epidémiologie : Principes et méthodes quantitatives. Editions Inserm Paris 193: p198-199).
Description of the cohort as a whole is extensive and is not particularly useful and pertinent to the study- if the authors want to include this, readers can generate this from the table and don’t need to have it stated in the text.

**Answer to reviewer:** It is important to provide a brief description of the overall cohort in a cohort study so that readers can compare this population to their own. We have shortened this section, as requested.

There is a large amount of data here, some of which is not relevant. What is the benefit to the study of stating the proportion of women that lived alone?

**Answer to reviewer:** Women who are pregnant and living alone are at greater risk of social and economic vulnerability and thus also theoretically at higher risk of some diseases (such as diabetes). This has been observed in French studies focusing on social and economic vulnerability. It turned out that there was no difference between our two groups but this information is interesting for obstetricians. It is relevant in non-randomized studies to search for differences in confounding and prognostic factors between the two groups. We have deleted Gynecological History from Table 2 to shorten it.

What is meant by gynaecological history – what history? How is this relevant? You would assume the control group to have a longer gestational age. Is there more information about the 25% or so that were induced in the control group – what was the upper limit for gestation for these inductions?

**Answer to reviewer:**
- For the gynecological history variable, as stated above, it has been removed from the table.
- The gestational age was indeed higher in the control than in the interventional group, as is inherent in the definition of the groups, and fortunately, for otherwise the study would raise methodological problems. We want to know if it might be useful to induce labor at 37-38 weeks +6 days in women with a fetus suspected of being large for gestational age. The hypothesis presented, which is explained in the article (page 12), is that labor should be induced early enough in these women to expect a benefit, for the child will not continue to gain weight. This has not always been tested rigorously in previous studies, which have compared women who have been induced either with those not induced, at all terms combined, or with an inadequate contrast between the two groups for term at induction (including the small randomized trials).

For the control group, inductions were performed after the term chosen for the experimental group (37-38 weeks +6 days) either because a disease or disorder appeared in mother or child or because term was reached.

National clinical practice guidelines were issued in France in 2008 about the induction of labor at term (but this practice is of long standing in France) (http://www.has-sante.fr/portail/upload/docs/application/pdf/declenchement_artificiel_du_travail-_synthese.pdf). These guidelines state that the risk of complications associated with post-term delivery mandates specific monitoring from the day term is reached. The following outline is recommended: dates should be treated as ± 1 day:

- If the women has not given birth by 41 weeks +6 days, fetal monitoring should take place every 48 hours.
- If delivery has not occurred by 41 weeks +6 days, induction of labor is recommended, possibly preceded by cervical ripening with prostaglandins.
It is possible to induce labor at 41 weeks + 0 day, on condition that the cervix is favorable and after informing the woman and obtaining her consent. This approach may be motivated by the impossibility of regular monitoring, the woman's request, or the needs of the organization of care.

A sentence has been added to page 7 to explain the reasons for induction for women in the group control.

**French practices are consistent with the currently available evidence** [Gülmezoglu AM1, Crowther CA, Middleton P, Heatley E. Induction of labour for improving birth outcomes for women at or beyond term. Cochrane Database Syst Rev. 2012 Jun 13;6:CD004945. doi: 10.1002/14651858.CD004945.pub3].


The distribution of term at induction in the control group is reported underneath Table 2, to avoid weighing down the text. Similarly, for the same reason, we have not added to the results section the causes of induction in the control group, which add no further information (planned or term = 38.8%; maternal disease = 11.8%; fetal disease = 20.9%; rupture of membrane before labor = 14.7% and other = 13.9%).

The type of operative vaginal delivery needs clarification. The rate of this type of delivery appears relatively high (indeed would be for the UK where is is around 10-15%).

**Answer to reviewer:** Beneath Table 3, we have described more precisely the details of the instrumental deliveries in the cohort, while we reported the content of the variable of operative vaginal delivery in the Table, for more clarity. We did this to increase the legibility of the tables. We included in the operative deliveries those using instruments and those involving obstetric maneuvers.

A publication from the Audipog database reported among singleton deliveries: 5.6% of instrumental deliveries with forceps; 3.9% with spatulas; 3.2% by vacuum extraction, 0.1% breech extraction, and 0.2% other maneuvers [Vendittelli et al. Réseau Sentinelle audipog 2004-2005. Partie 1 : résultats des principaux indicateurs périnatals. Gynecol Obstet Fertilite 2008;36:1091-1100].


Surely the increased spinals in the control group reflects the increased no of CS in this group?

Answer to reviewer: Of course. We have modified the sentence slightly to clarify it.

The paragraph beginning: ‘We identified the following confounding factors’ makes very little sense to me and needs to be re-written. How did the authors define ‘confounding’ factors in this analysis (if this is what it was) – do you mean these were potential co-variates?

Answer to reviewer: In epidemiology, there are both known confounding factors and others for which one must search. The method was clearly explained on page 8 in the statistical analysis. This work has been reviewed by 2 statisticians. From an epidemiologic perspective, a confounding factor is a third factor that is simultaneously associated with the exposure under study and the disease being studied (the endpoint). We do not think that this does not require a more detailed definition in the text.

The sentence beginning ‘Nor did the adjusted risk of episiotomy…’ in the paragraph on perineal lesions should come before the subgroup analysis for primips/ multips, as this appears to refer to all the patients and is not subdivided.

Answer to reviewer: We have moved the sentence to the beginning of the paragraph, but we note that episiotomy is included among perineal lesions.

This sentence: ‘Accordingly, the control group included more cesareans (31.2% vs. 21.6%) (p=.005)’ is at odds with ‘The crude risk of a cesarean during labor in the induction group was no higher than in the control group (RR=1.16; 95%CI: 0.88-1.53).’ This sentence is then followed by ‘after adjustment’ – adjustment for what? This analysis needs clarification.

Answer to reviewer: We have corrected the sentence. The adjustment factors have been moved to the bottom of the two tables to disencumber the text. They are also listed on page 10.

What did the authors mean by ‘traumatic neonatal lesions’?

Answer to reviewer: This endpoint was described below Table 5. We have added its definition to page 8, in the methods section.

Are estimated fetal weights plotted on customized growth charts in France (is this routine practice) or is this something that the authors of the study did? Or were the clinicians acting on birth weight alone?

Answer to reviewer: In France, clinicians generally use paper curves or an automatic calculation from either their computerized files or our web site (http://www.audipog.net/courbes_morpho.php). Our curves are very widely used. The method is as follows: The mean weight and its standard deviation are calculated for each gestational age (in weeks of gestation) for boys and for girls. The normality of the distribution is verified for each subsample. To model the mean weight and its standard deviation, a weighted polynomial regression is performed for each of these two measures. The Z score can be calculated easily from these models.  
With OW the observed weight, PW the predicted weight and PSD the predicted standard
deviation:
Z score = (OW – PW)/PSD.
Centile = (Φ(Z score) * 100) where Φ is the distribution function of the standard normal distribution.

The methodology is the same for length and head circumference at birth.
The Audipog website is listed in the references and is available in English, so that the reader can obtain further details. This is explained on page 7.

Besides these curves, we also have individual curves for each child (customized birth weight curves). These are used less often by clinicians in France and are not relevant to this study.
They involve modeling birth weight and length while taking into account gestational age, sex, the child’s birth rank, and the mother’s height and weight. This modeling makes it possible to calculate individual cutoff points below which a child must be considered to have experienced fetal growth restriction (assessed by weight or height or both).

Please clarify: ‘threshold > 97th percentile to match the definition of the international small-for-gestational age advisory board consensus statement of 2001’ – do you mean that SGA is defined as less than 3rd centile? It is not really ‘matching’ as this implies the same.

**Answer to reviewer:** We have corrected the sentence on page 12, to make it clearer and more accurate.

Remove the exclamation mark from p13

**Answer to reviewer:** We have deleted it.

Can the authors shed any light on differences seen between studies, for CS rates for example?

**Answer to reviewer:** Not really.

The discussion does not include anything about bias inherent in retrospective Analysis

**Answer to reviewer:** Yes, it did, on the previous p. 15, currently page 15. We have also included the comments of reviewer 2 and expanded this discussion. We also added one reference.

Reduction/ no change in CS rates in induction groups have been observed in studies comparing expectant management vs induction for medical problems – eg hypertension, diabetes studies, study of Sarah Stock et al. I think this study is in general agreement with this – but this is not discussed.

**Answer to reviewer:** We have modified the discussion on page 13 and added 2 new references.

« Wood S, Cooper S, Ross S. Does induction of labour increase the risk of caesarean section? A systematic review and meta-analysis of trials in women with intact membranes. BJOG 2013; DOI:10.1111/1471-0528-12328.” These authors found that induction of labor in women with intact membranes reduced the risk of cesarean section.
The study by Stock SJ et al (BMJ 2012) used an unselected population database linked to a database of neonatal outcomes. The principal problem in this study is that the Scottish database does not mention the indication for the induction of labor and that, in the absence of any maternal or fetal pathology, the induction of labor was considered to be elective. Neither of these points applies in our study. They stressed that the cesarean rate varied according to term, but the reduction in the number of postpartum hemorrhages was consistently observed.

There is also no discussion about the fact that women in the 37-38 week induction group seemed to have another indication for induction – eg polyhydramnios and the impact this has on interpretation of the results. The text generally is quite long and repetitive at times, and could be written much more concisely.

**Answer to reviewer:** The search for confounding and prognostic factors does weigh down the results section, but also shows the methodological seriousness of the work. Without identifying these variables, we cannot take them into account in a multivariate analysis, as we very carefully did. Our adjusted RRs are therefore not biased. Our analyses took into account the differences observed between the groups. Now, as we have note in the discussion, it is possible that we did not identify (because they were unknown or unavailable in our database) other factors that could have biased our RR.

Assessing a medical care policy (here, induction for large-for-gestation-age fetuses) does not prevent a related policy applying to a different situation (here, that the control group undergo induction later on for another reason (medical or not). This approach is that applied to pragmatic intention-to-treat randomized trials on the subject.

**Minor compulsory revisions**
**Change:** ‘reduce the onset of perineal tears’ to ‘reduce the occurrence of…’

**Answer to reviewer:** We have made this correction.

**Discretionary revisions**
- Would be usual to use the terminology retrospective, rather than historical

**Answer to reviewer:** We use the term retrospective, but prefer to describe this as a historical cohort study, following the Dictionary of Epidemiology, 2nd edition, edited by John Last for the International Epidemiology Association, which describes a historical cohort study as “a cohort study conducted by reconstructing data about persons at a time or times in the past. This method uses existing records about the health or other relevant aspects of a population as it was at some time in the past and determines the current or subsequent status of members of this population with respect to the condition of interest. Different levels of past exposure to risk factor(s) of interest must be identifiable for different subsets of the population.” Samet & Munoz point out that historical, retrospective, and non-concurrent cohorts are all used to describe this and recommend that “Any lingering debate with regard to terminology should be set aside to avoid unneeded confusion.” Because our study specifically avoids a major pitfall of retrospective studies, specifically in that in utero weight is estimated during pregnancy, before the child's birth and thus before birth weight is known, use of the term might confuse and mislead readers about this important point, which constitutes the originality of this work.
Reviewer 2: Bukola Fawole

**Reviewer's report:**
The Authors investigated an historical cohort of women obtained from the French database for 233 health facilities between 1994 and 2008 with the goal of determining whether a policy of elective induction of labour between 37 and 39 weeks for women with ultrasound suggestion of large for gestational age (LGA) fetuses (i.e. > 97th centile) would reduce perineal tears. They identified 199 such cases and compared with 2878 controls, also LGA, that were delivered within the same gestational period but without induction of labour. Authors reported that of all cases suspected of LGA, only 26.9% had birthweight > 97th centile.

Major compulsory revisions
Authors need to indicate for 2878 controls, what proportions had spontaneous onset of labour and elective caesarean section.

**Answer to reviewer:** In Table 2, we inserted the criterion "mode of onset of labor including induction of labor ≥ 39 weeks (and not induction regardless of term because our factor of exposure is "induction of labor at 37-38 weeks +6d"), spontaneous labor" and "elective cesarean". Consequently, the table no longer fits on a single page.

The controls were reported to have delivered within > 37 and < 38 weeks 6 days, yet Table 1 reports that 25.7% had induction of labour at > 39 weeks. Such cases obviously did not meet eligibility criteria.

**Answer to reviewer:** As requested by reviewer 1, we have reformulated both the text and the tables, to make them easier to understand.

“Among the fetuses suspected of LGA status before birth, we compared those whose mothers’ labor was induced between ≥ 37 weeks and ≤ 38 weeks+6 days (n=199) to those with expectant obstetrical management during this time period (thus considered the unexposed group)(n=2878).”

That is, we compared a policy of routine induction of labor between ≥ 37 weeks and ≤ 38 weeks+6 days among fetuses suspected of macrosomia in utero to a policy of routine non-induction between ≥ 37 weeks and ≤ 38 weeks+6 days, among the same population. It is therefore normal that the control group would undergo spontaneous deliveries and induction at a later term, as explained above, either because the pregnancy reached term or because a late-term pregnancy disease appeared. This is precisely the originality of this work, to have a
pragmatic intention-to-treat analysis, from a database that includes the variable of suspected macrosomian before birth weight is known. Inductions of labor after 39 weeks are therefore naturally expected; the control group gave birth at various terms (they most especially did not, on the other hand, have labor induced between > 37 and < 38 weeks \( \pm 6 \) days (see also page 12, in the discussion).

The study period was between 1994 and 2008. The skills of the Sonographers/Sonologists must have been variable and this also constitutes a weakness of the study.

**Answer to reviewer:** We have added this concept to the discussion. In France, the training of ultrasonographers has been standardized since before 1994. It is more the performance of the ultrasound instruments that has advanced (see page 14). We described page 11 that “Fetal weight is routinely estimated in France during the last fetal ultrasound, performed by a well-trained professional ».

Minor essential revisions
On page 6, lines 2 and 3, Authors wrote '......might reduce onset of perineal tears....'. Did they intend to write incidence?

**Answer to reviewer :** We would like to say « occurrence ». 