PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Temporal trends in neonatal mortality and morbidity following spontaneous and clinician-initiated preterm birth in Washington State, U.S.A.: a population-based study
AUTHORS	Richter, Lindsay; Ting, Joseph; Muraca, Giulia; Synnes, Anne; lim, Kenneth; Lisonkova, Sarka

VERSION 1 – REVIEW

REVIEWER	Elsa Lorthe Inserm UMR 1153, Obstetrical, Perinatal and Pediatric Epidemiology Research Team (Epopé), Center for Epidemiology and Statistics Sorbonne Paris Cité, Département Hospitalo-Universitaire Risks in pregnancy, Paris Descartes University, France
REVIEW RETURNED	07-May-2018

GENERAL COMMENTS	The authors aimed at describing temporal trends in PTB subtypes and related rates of neonatal mortality and morbidity within two databases in Washington State from 2004 to 2013. The manuscript is well-written. However, I recommend consideration of the following revisions. Abstract: Objective: State precisely where (in which country) and when the PTB rate has declined. Do you refer to the overall PTB rate or to the PTB rate among singleton births only (as multiple pregnancies can have a great impact on PTB rate)? Outcome measures: Why did you use categories of GA (24-27, 28-31, 32-33, 34-36) instead of weeks of GA (24, 25, 26 etc)? Provide a clear definition of the composite outcome “any severe neonatal morbidity”, without “etc” (for instance, is neonatal sepsis included in this definition as the conclusion suggests?). Results: please provide a 95% CI for the overall neonatal mortality rate, and p-value for the increase in the composite outcome. Stata clearly what is the reference when reporting AORs. Conclusions: The sentence “…with respect to prevention of PTB” is not supported by data and should be removed. Introduction: Clear. I disagree with the “following PPROM” part of the sentence “preterm infants born to women with spontaneous onset of labor have a better prognosis than infants born following PPROM or iatrogenic intervention”, as recent publications did not report any difference in neonatal prognosis when comparing spontaneous PTB and PPROM (Delorme 2016, Torchin 2017, Chevallier 2017…).

	Methods: Please provide the precise numbers of infants excluded because of birth<24w, >45w etc (+ also in the flow chart). Why did you exclude stillbirths, that can be seen as competitive events with neonatal mortality and morbidity? Can this exclusion impact the evolution of the composite outcome neonatal death/severe morbidity? Please discuss this point in the comment section. There is an inconsistency between the exclusion of stillbirths from the study population (page 6), and the assessment of temporal trends in stillbirths (page 7). Please clarify. Please explain which data were used to classify PTB subtypes (directly reported in databases? Any validation of these variables?) Which curve was used to classify infants as SGA <10th percentile? Why did you use categories of GA (24-27, 28-31, 32-33, 34-36) instead of weeks of GA (24, 25, 26 etc)? How did you choose the confounders to adjust for in multivariate models? Please, clarify if “missing values <3%” refers to each covariate (in this case, report the number of patients included in the complete-case analysis) or to the total frequency of patients with at least 1 missing data (and thus not included in complete-cases analyses). Which p-value was considered significant (do not provide this information only in the tables)? I assume that one should choose a p-value lower than the usual 0.05, which inappropriate with such a sample size and multiple comparisons. Results Table 1 (+ methods): consider reporting a global p-value when comparing variables with more than 2 categories (and not only p-values by category) Table 1: please explain why years are classified as 2004-2006, 2007-2010 and 2011-2013, and if the p-value (Cochran-Armitage test) compares these 3 periods or all the years. Please harmonize the choice of the test with data provided in the tables. Table 1: I assume that on page 12 you provide p-values corresponding to percentages comparisons while you report effectives in the table (for instance 489, 746, 598 at 24-27 w). Please clarify, and report % in the table if appropriate. Table 1: It appears that the proportion of spontaneous labor remains stable within all the sub-categories of gestational age. Can the decrease in the frequency of PPROM (or the increase in the frequency of iatrogenic deliveries) be related to a difference in coding data? Please discuss if any misclassification is possible (for instance PPROM classified as iatrogenic delivery if a cesarean section was performed before labor because of abnormal FHR or chorioamnionitis)?

	Figures 1 & 2: do you have any explanation for the increase in the overall PTB rate in 2012? Is it related to a change in data collection? If so, the subsequent decline could be an artificial consequence of this increase. Table 2: Why are the outcomes rates not reported for the period 2007-2010? Tables 2, 3 & 4: Why do you report rates per periods but adjusted odds ratio per 1-year change? Is it the most appropriate measure if the trend is non linear? Discussion: How do you explain the differences between your findings and those reported by Gyamfi-Bannerman? What are the hypothesis to explain the increased mortality/severe morbidity in late preterm infants according to declines in SPTB and PPROM and increase in iatrogenic deliveries? Can the increased proportion of SGA infants be responsible for increased iatrogenic deliveries and increased adverse outcomes? Page 19: Giving explanations for variations on PTB rates is challenging as not supported by data reported here (for instance do not conclude on the role of cerclage or multiple births as they were not included in this analysis). Please comment on the potential impact of the 116886 + 2549 + 14503 infants excluded on the reported trends. What do you mean by “iatrogenic TOP”? Page 20: “which points to possible common causes relatively independent of delivery type”: the physiopathology of neonatal sepsis could also be different by PTB subtype (related to chorioamnionitis in PPROM and to SGA in iatrogenic deliveries for instance), or be related to the approximative definition used in this article. Page 21: the most likely misclassification would be the complicated cases of PPROM classified as iatrogenic deliveries, I doubt that this can be considered non-differential. Elaborate why you think that errors and omissions can result in underestimation of temporal trends.

REVIEWER	Howard L. Sobel World Health Organization, Western Pacific Regional Office
REVIEW RETURNED	16-Jul-2018

GENERAL COMMENTS	This is a useful manuscript with a comprehensive analysis. However, it should be rewritten with the target audience in mind. Specific comments: 1. As this is an international medical journal, the global readership will be interested in implications beyond the need for further research. What should the public health department, hospital

	management and clinicians be considering on the basis of this findings plus those of related studies? This should be included in the abstract and discussion. 2. The readership may not know about Washington State (e.g., vs Washington D.C.). The introduction should include a description of the state including population size, how it fares in preterm births and mortality compared with other states (e.g., where does it rank in the US). Likewise, the international readership would probably be interested how the US and the state measure up to other high-income (and maybe high-middle income) countries. The discussion would then benefit from putting the findings into broader context (neighboring states, US as a whole and global). 3. The readership will probably confuse iatrogenic and unnecessary procedures leading to preterm birth. In common usage in my areas of work, the two are almost synonymous. Obviously the critical importance is that unnecessary iatrogenic procedures that result in high risk of adverse outcomes are an immediate target for intense monitoring and enforcement of standards. However, the intro describes it as "iatrogenic (medically-indicated)". This immediately prompts several questions: Are all iatrogenic preterm births truly medically indicated? How can the reviewers be sure? What safeguards are in place? (e.g., independent validation of records.) How did Washington State get around false reporting? (i.e., it being written as indicated when in fact it was not.) In the past two years, I have personal contacts with two persons who had unnecessary cesarean, one with a failed induction where the lie was require surgical intervention (admittedly not from Washington State). Bottom line, more description is needed that iatrogenic procedures were necessary. Otherwise, the proportions necessary or unnecessary should be included and reanalysis of proportion of adverse outcomes that could be prevented if unnecessary iatrogenic preterm births eliminated as well as the major implications of the unnecessary ones in abstract, discussion and conclusions. If these are not clearly available, this is a major limitation and should be stated as such. (This goes well beyond line 31-32 in the paragraph with limitations). 4. "The completeness and accuracy of these databases was monitored by the Washington State Department of Health with annual assessments and consistency checks.17-19" The reviewer believes more description of the validation mechanisms. Has Washington State solved the age-old problem of inaccurate completion of death certificates and other reporting forms some of which were used in the study? This includes intentional and unintentional inaccurate reporting. If truly validated, then the accuracy (not the completeness) should be stated in the manuscript. 5. The non-statistical specialist readership (and reviewer) may not understand why three periods of three years were included instead of an annual rate. This should be spelled out. If this is a standard, it should be explained and referenced 6. The STROBE Statement is a very useful tool for both authors and reviewers. The reviewer recommends in addition to stating the place that may answer the statement, to put in quotes the actual statement.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1

Abstract

1. Objective: State precisely where (in which country) and when the PTB rate has declined. Do you refer to the overall PTB rate or to the PTB rate among singleton births only (as multiple pregnancies can have a great impact on PTB rate)?

RESPONSE: The preterm birth rate has declined in the United States since 2006, after peaking at 12.8% (Gyamfi-Bannerman, Obstet Gynecol 2014). The temporal trend in preterm births has continued to decline, and in 2013 (the most recent year included in this study), the preterm birth rate was reported to be 11.4%, which was the lowest preterm birth rate in the United States since prior to 2000 (Frey, Seminars in Fetal and Neonatal Medicine 2016). The reported preterm birth rate refers to the overall preterm birth rate including all births – singletons and multiples. Multiple pregnancies do have a great impact on the preterm birth rate. In 2013, the preterm birth rate in the United States among singletons only was 9.7%, among twins was 56.6%, and among triplets was 93.1% (National Vital Statistics Reports, Vol 64, No 1, January 15, 2015). Understandably, singletons and multiples should be analyzed separately, and this study focuses on singleton preterm births. The preterm birth rate among singleton births declined from 2005 to 2012, with the largest decline (15.8%) among singleton late (34-36 weeks) preterm births (Gyamfi-Bannerman, Obstet Gynecol 2014). The most recent data show a decline between 2006 and 2015 (https://www.marchofdimes.org/peristats/tools/ReportFiles/PrematureBirth/2017/rc/pdf/PrematureBirthReportCard-UnitedStates-2017.pdf).

The following changes have been made to the manuscript abstract:

Page 2: “After a decade of increase, the preterm birth (PTB) rate has declined in the United States between 2006 and 2015 with the largest decline at late preterm (34-36 weeks). We described concomitant changes in gestational age-specific rates of neonatal mortality and morbidity following spontaneous and clinician-initiated PTB among singleton infants.”

“The temporal decline in late PTB among singleton infants was associated with increased mortality among late preterm infants following clinician-initiated delivery and increased combined mortality or severe morbidity among all late preterm infants, mainly due to increased rate of sepsis.”

2. Outcome measures: (a) Why did you use categories of GA (24-27, 28-31, 32-33, 34-36) instead of weeks of GA (24, 25, 26 etc)? (b) Provide a clear definition of the composite outcome “any severe neonatal morbidity”, without “etc” (for instance, is neonatal sepsis included in this definition as the conclusion suggests?).

RESPONSE: (a) We used well defined and clinically meaningful categories of gestational age that are extensively used in the literature. The categories are subdivided into births that occur at less than 28 weeks’ (extreme prematurity), 28-31 weeks’ (severe prematurity), 32-33 weeks’ (moderate prematurity), and 34-36 weeks’ (near term/ late prematurity) (Goldenberg, Lancet 2008). We chose to use the categories of gestational age instead of individual weeks of gestation because some outcomes are rare and the rates per week of gestation would be statistically unstable. The defined categories still allow clinically relevant conclusions to be drawn.

(b) In the abstract, all eight individual morbidities included in the composite outcome were not explicitly named due to word limitations. However, the exact components of the composite are listed in the Methods section and defined in the appendix (Appendix Table 1).

The abstract was revised as follows:

Page 2: “The primary outcomes were neonatal mortality and a composite outcome including death or severe neonatal morbidity.”

3. Results: (a) Please provide a 95% CI for the overall neonatal mortality rate, and p-value for the increase in the composite outcome. (b) State clearly what is the reference when reporting AORs.

RESPONSE: (a) The sentence describing the overall neonatal mortality rate was revised to include the 95% CI as follows:
Page 2: “Overall neonatal mortality remained unchanged (1.3%; AOR 0.99, CI 0.95-1.02)”;
and the sentence describing the increase in the composite outcome was revised to include the 95% CI (instead of p-value for consistency) as follows:
“overall rate of the composite outcome increased (from 7.9% to 11.9%; AOR 1.06, CI 1.05-1.08).”
(b) The AORs express the average annual change in the odds of the outcome; therefore, the reference was the prior calendar year that was modelled as a continuous variable. The AOR is described in the outcome measures section of the abstract as follows:
Page 2: “Logistic regression yielded adjusted odds ratios (AOR) per 1-year change in outcome…”.

4. Conclusions: The sentence “…with respect to prevention of PTB” is not supported by data and should be removed.
RESPONSE: The sentence was removed and the conclusion statement was revised as follows: Page 2: “The temporal decline in late PTB was associated with increased mortality among late preterm infants born following clinician-initiated delivery and increased combined mortality or severe morbidity among all late preterm infants, mainly due to increased rate of sepsis.”

Introduction
5. I disagree with the “following PPROM” part of the sentence “preterm infants born to women with spontaneous onset of labor have a better prognosis than infants born following PPROM or iatrogenic intervention”, as recent publications did not report any difference in neonatal prognosis when comparing spontaneous PTB and PPROM (Delorme 2016, Torchin 2017, Chevallier 2017…).
RESPONSE: The PPROM part of the sentence was removed, and the sentence was revised as follows:
Page 4: “Preterm infants born to women with spontaneous onset of labor have a better prognosis than infants born following clinician-initiated delivery”.

Methods
6. Please provide the precise numbers of infants excluded because of birth<24w, >45w etc (+ also in the flow chart).
RESPONSE:
We added more detailed description of excluded records in the flow chart (Appendix Figure 2).
TOTAL 871649 percent
Multiples 24991 2.87
Gestation <24 weeks 3165 0.36
Gestation >45 weeks 27 0.00
Gestation week missing 3758 0.43
Out-of-state/out-of-hospital/unmatched births 84945 9.75
Overall excluded 116886 13.41

7. Why did you exclude stillbirths, that can be seen as competitive events with neonatal mortality and morbidity? Can this exclusion impact the evolution of the composite outcome neonatal death/severe morbidity? Please discuss this point in the comment section.
RESPONSE: The temporal trend in the rate of stillbirth among all hospital births was examined and is now described in the results section as follows:
Page 10: “The rate of stillbirth increased slightly from 3.2 per 1000 total births in 2004-2006 to 3.7 in 2011-2013 (p=0.002). Stillbirth rates increased at 24-27 weeks (from 0.7 to 1.0 per 1000 fetuses-at-risk [FAR], p=0.003), and at 28-31 weeks’ gestation (from 0.4 to 0.7 per 1000 FAR, p=0.002; Appendix Figure 3).”
After the assessment of temporal trends in stillbirths, the 2549 stillbirths were excluded because our

objective was to describe neonatal outcomes following clinician-initiated vs spontaneous delivery at preterm. Delivery of stillborn infant is commonly iatrogenic (as the labor is induced within one or two days after stillbirth, unless spontaneous birth occurs to deliver a dead fetus). As such, mode of delivery is irrelevant. The effect of changes in clinician-initiated vs spontaneous PTB rates on stillbirth rates have to be examined separately using ecological design (e.g., positive or negative correlation with the changes in clinician-initiated PTB rates). This study aims to describe temporal changes in neonatal prognosis after clinician-initiated vs spontaneous PTB, hence the denominator includes live births only. The results show that stillbirth rate increased at 24-31 weeks, which may have led to a lower rates in neonatal mortality and morbidity composite outcome over time than we would have expected, had these stillborn fetuses were born alive potentially seriously compromised.
We included this point in the Discussion section:
Page 20:" However, the vast majority of iatrogenic pregnancy terminations is likely to occur prior to 24 weeks gestation; terminations beyond 23 weeks would be included as stillbirths in this study. Temporal changes in gestational age-specific stillbirth rates showed small increases in stillbirth rates at 24-27 weeks and 28-31 weeks, which augments the upward trend in adverse neonatal outcome (mortality or severe morbidity) at 28-31 weeks gestation."

8. There is an inconsistency between the exclusion of stillbirths from the study population (page 6), and the assessment of temporal trends in stillbirths (page 7). Please clarify.
RESPONSE: The total study population including all singleton hospital births did include stillbirths and these were analyzed separately from the live birth study population. The exclusion criteria have been clarified in the Methods section as follows:
Page 6: "We excluded infants born less than 24 weeks' and greater than 45 weeks' gestation, and those with missing data on gestational age from the overall study population. After analysis of temporal trends in stillbirth, we excluded stillborn infants and those with missing mode of delivery to limit the analyses of neonatal outcomes following various types of preterm birth to live births only."
See also answer to comment #7.

9. Please explain which data were used to classify PTB subtypes (directly reported in databases? Any validation of these variables?)
RESPONSE: We classified preterm birth subtypes using information included in the Washington State Birth Events Record Database (BERD); and we now attached Washington State Birth Filling Form as a Supplemental file to the manuscript (Appendix 1). We first categorized births into PPROM category (item #64 in the BERD data collection form), then clinician-initiated delivery (information on induction – item #65, and cesarean delivery – except cesarean with a trial of labor – item #62), and spontaneous onset of labor (all other births, also item #64).
The following information was added to the manuscript:
Page 6: "Preterm birth subtypes were categorized using the following algorithm: (1) first, spontaneous preterm births following PPROM (>12 hours); (2) second, clinician-initiated preterm births following labor induction or cesarean delivery without labor; and (3) third, all other births were classified as spontaneous preterm births following spontaneous labor onset with intact membranes (Appendix 1, items no 62, 64, 65)."
We added more details about accuracy and validation of the dataset to the Methods section:
Page 6: "The completeness and accuracy of these databases was monitored by the Washington State Department of Health with annual assessments and consistency checks. (16–18) Records flagged with inconsistent or out-of-range entries were addressed systematically through hospital review and correction. The frequency of diagnostic and procedure codes was monitored in annual reports.(18) Previous validation studies of the linked dataset showed that the positive and negative predictive values (PPV and NPV) for delivery characteristics were above 80% and 98%, respectively;(19,20) for example, labor induction had PPV 89.0% and NPV 94.5%.(20)"

10. Which curve was used to classify infants as SGA <10th percentile?
RESPONSE: SGA infants were classified using the following reference: .
"Alexander G, Himes J, Kaufman R, et al. A United States reference for fetal growth.
Obstet Gynecol 1996;87:163-8."
This reference was added to the revised manuscript.

11. Why did you use categories of GA (24-27, 28-31, 32-33, 34-36) instead of weeks of GA (24, 25, 26 etc)?
RESPONSE: Please see the response to comment #2.

12. How did you choose the confounders to adjust for in multivariate models?
RESPONSE: The confounders chosen to adjust for in the multivariate models (i.e., maternal age, pre-pregnancy BMI, race, education, smoking, marital status, parity, chronic hypertension, pre-pregnancy diabetes, assisted conception, health insurance provider, gestational age, SGA infant, sex and congenital anomalies) because their temporal changes were found to be significant in the univariate analysis and/or they were known risk factors for adverse neonatal outcomes.
We added the following to the Methods section:
Page 8: "Logistic regression was used to assess temporal trends in adverse neonatal outcomes adjusted for temporal changes in risk factors that may have changed over the study period:…"

13. Please, clarify if "missing values <3%" refers to each covariate (in this case, report the number of patients included in the complete-case analysis) or to the total frequency of patients with at least 1 missing data (and thus not included in complete-cases analyses).
RESPONSE: Complete case analysis included 93% of preterm births, we clarified this in the Methods section
Page 9: "Other missing values were <3.0% of the total, and the complete case multivariable analysis excluded 7.0% of preterm births."

14. Which p-value was considered significant (do not provide this information only in the tables)? I assume that one should choose a p-value lower than the usual 0.05, which inappropriate with such a sample size and multiple comparisons.
RESPONSE: We refrained from the emphasis on significant p-value cutoffs. This position is well supported by statisticians. The P-values for trend are presented for descriptive purposes to provide information from the statistical perspective, whether the temporal trend is present besides the effect of year-to-year variability in the rates. We clarified this issue in the Methods as follows:
Page 9: "All p-values are reported as recommended by the American Statistical Association (24)."

Results
15. Table 1 (+ methods): consider reporting a global p-value when comparing variables with more than 2 categories (and not only p-values by category)
RESPONSE: Table 1 has been revised to include a global p-value when comparing the variables with more than 2 categories instead of by individual categories.

16. Table 1: please explain why years are classified as 2004-2006, 2007-2010 and 2011-2013, and if the p-value (Cochran-Armitage test) compares these 3 periods or all the years. Please harmonize the choice of the test with data provided in the tables.
RESPONSE: In Table 1, the middle period (2007-10) was removed from the table in order to describe the comparison between the beginning and the end of our 10-year study period (2004-06 and 2011-13). A new column was added to describe the total numbers for each demographic and clinical characteristic for all 10 years (2004-2013). The table was also revised to include a p-value for the Chi-square test comparing Period 1 (2004-06) and Period 2 (2011-13) rather than a p-value for the

Cochran-Armitage test for temporal trend over the entire study period.

17. Table 1: I assume that on page 12 you provide p-values corresponding to percentages comparisons while you report effectives in the table (for instance 489, 746, 598 at 24-27 w). Please clarify, and report % in the table if appropriate.
RESPONSE: This section of Table 1 was removed and added to a new appendix table (Appendix Table 2) titled “Preterm live births by gestational age categories and clinical subtype.” The table 1 was also revised (see response to the comment #16).

18. Table 1: It appears that the proportion of spontaneous labor remains stable within all the sub-categories of gestational age. Can the decrease in the frequency of PPROM (or the increase in the frequency of iatrogenic deliveries) be related to a difference in coding data? Please discuss if any misclassification is possible (for instance PPROM classified as iatrogenic delivery if a cesarean section was performed before labor because of abnormal FHR or chorioamnionitis)?
RESPONSE: It is unlikely that this is due to a difference in coding data. PPROM was abstracted from medical charts separately from the mode of delivery. When PROM was present and a caesarean section was performed, this birth would still be categorized as PPROM preterm birth and not as clinician-initiated delivery. PPROM diagnosis criteria and data collection process were not changed during the study period.
We clarified the algorithm for PTB subtype in the Methods:
Page 6: “Preterm birth subtypes were categorized using the following algorithm: (1) first, spontaneous preterm births following PPROM (>12 hours); (2) second, clinician-initiated preterm births following labor induction or cesarean delivery without labor; and (3) third, all other births were classified as spontaneous preterm births following spontaneous labor onset with intact membranes (Appendix 1, items no 62, 64, 65).”

19. Figures 1 & 2: do you have any explanation for the increase in the overall PTB rate in 2012? Is it related to a change in data collection? If so, the subsequent decline could be an artificial consequence of this increase.
RESPONSE: There is not apparent explanation for the increase in the overall preterm birth rate in 2012. The increase appears to be present in clinician-initiated and spontaneous labor subtypes but not among PPROM. Diagnosis criteria and data collection process were not changed during the study period. However, we cannot rule out that changes in physicians’ preferences for mode of delivery (e.g., trial of labor before cesarean delivery) were responsible for this one-year excess variation in the overall temporal trend.
We added the following to the Discussion:
Page 20: “Fifth, the data sources had detailed information on mode of delivery that allowed accurate categorization of preterm birth subtypes; however, this categorization may have overestimated the proportion of deliveries following PPROM.(43) Data collection had not changed over the study period, however, changes in physician’s preferences for specific mode of delivery (e.g., trial of labor before cesarean delivery) may be responsible for year-to-year fluctuation in temporal trends in preterm birth subtypes.”

20. Table 2: Why are the outcomes rates not reported for the period 2007-2010?
RESPONSE: Table 2 has been revised such that all tables are consistent with respect to their formatting. This period was reported in other tables for descriptive purposes but has been removed after reviewer’s comments.

21. Tables 2, 3 & 4: Why do you report rates per periods but adjusted odds ratio per 1-year change? Is it the most appropriate measure if the trend is non linear?
RESPONSE: The rates by period were reported for descriptive purposes, whereas, for analytical

purposes, we used statistical methods to assess the trend as approximately linear with the AOR expressing average 1-year change in the outcomes. These tables have been revised to include only the rates for the beginning and the end of the 10-year study period (2004-06 and 2011-13), a crude rate ratio between these two periods, and lastly, the AOR per 1-year change in outcome in order to statistically evaluate the temporal trends over entire study period. Considering entire study period with respect to temporal trend is more robust from the statistical point of view. Our aim was to detect overall trends (upward or downward) over the 10-year period. After examining the graphs, we did not find any strong indications of exponential or other higher-level polynomial trends and therefore we proceeded with analysis using logistic regression. It was not our aim to model the minor changes in outcome rates over the years (for example, using restricted cubic splines).

Discussion

22. How do you explain the differences between your findings and those reported by Gyamfi-Bannerman?

RESPONSE: Our findings and those reported by Gyamfi-Bannerman et al. are similar, the only discrepancy is in the trends in clinician-initiated preterm birth. Both studies (Gyamfi-Bannerman et al. and our study) included singleton live births and a similar study period (2004-2013 vs. 2005-2012), however, the study population were different: Gyamfi-Bannerman's study included all of the United States, whereas our study was limited to Washington State only. The Washington State population may vary from the total United States population in terms of demographic composition and risk factors of preterm birth. However, the Washington State birth certificate data have more detailed information on PPROM, mode of delivery and other characteristics, e.g., BMI. Gyamfi-Bannerman's findings of concurrent declines in both spontaneous and clinician-initiated preterm births are different from our findings. We considered PPROM and spontaneous preterm births separately, and found a temporal decline in the rates of both. However, we found a slight increase in clinician-initiated preterm births. These differences may be due to the inherent population differences or the different definitions of preterm birth subtypes. Further, Gyamfi-Bannerman's study focused on preterm birth trends, while our study focused mainly on neonatal outcomes following various types of preterm birth.

We included comparisons with Gyamfi-Bannerman et al.'s study in the Discussion:

Page 17: "A recent study by Gyamfi-Bannerman et al. showed a decline in both clinician-initiated and spontaneous preterm birth rates between 2005 and 2012.(10) Our study provides more detailed information on preterm birth categories and describes temporal trends in neonatal outcomes adjusted for changes in important risk factors."

23. What are the hypothesis to explain the increased mortality/severe morbidity in late preterm infants according to declines in SPTB and PPROM and increase in iatrogenic deliveries?

RESPONSE: The possible reasons for increased mortality/severe morbidity in late preterm infants given the decline in spontaneous preterm births and increased clinician-initiated preterm births are speculative. One hypothesis includes increasing proportion of infants born via clinician-initiated intervention at late preterm due to worsening in-utero condition of the fetus with subsequent increased risk of adverse outcome (for a newborn). Such trend may results from temporal increase in maternal risk factors that we did not adjust for (e.g., some co-morbidities such as asthma). As these potential indications for clinician-initiated delivery increase in the obstetric population, more infants are born preterm following clinician-initiated interventions that have adverse outcomes. However, these are hypotheses only and need to be examined further in large-scale, in-depth studies of neonatal outcomes. We adjusted for temporal changes in SGA, therefore SGA is not likely responsible for increased rates of adverse outcomes at late preterm in our study.

We added the following text to the Discussion:

Page 18: "However, at late preterm, declines in spontaneous and PPROM birth and increases in clinician-initiated delivery were associated with increased rates of mortality/severe morbidity. This may be due temporal increase in maternal chronic morbid conditions that we did not adjust for in our study,

for example, asthma, autoimmune conditions, or respiratory morbidity."

24. Can the increased proportion of SGA infants be responsible for increased iatrogenic deliveries and increased adverse outcomes?
RESPONSE: It is possible that the increased proportion of SGA infants may be responsible for the increased clinician-initiated deliveries. IUGR is an indication for iatrogenic delivery (for labor induction), especially close to term when the risks of prematurity become lower than consequences of adverse in-utero conditions leading to growth restriction and possible fetal demise. See also responses to the comment #23.

25. Page 19: Giving explanations for variations on PTB rates is challenging as not supported by data reported here (for instance do not conclude on the role of cerclage or multiple births as they were not included in this analysis).
RESPONSE: The sentence about the role of cerclage or multiple births had been removed.

26. Please comment on the potential impact of the 116886 + 2549 + 14503 infants excluded on the reported trends.
RESPONSE:
We added this limitation to the Discussion:
Page 21: "In addition, singleton infants excluded due to out-of-hospital delivery or missing values may have impacted our results, however, non-hospital births are more likely to be term deliveries without complications requiring hospitalization."

27. What do you mean by "iatrogenic TOP"?
RESPONSE: This was a typo, now corrected as "clinician-initiated PTB."

28. Page 20: "which points to possible common causes relatively independent of delivery type": the physiopathology of neonatal sepsis could also be different by PTB subtype (related to chorioamnionitis in PPROM and to SGA in iatrogenic deliveries for instance), or be related to the approximative definition used in this article.
RESPONSE: We included this comment in the Discussion as follows:
Page 18: "However, the pathology of neonatal sepsis can vary by preterm birth subtype (for example, originating from the effects of chorioamnionitis in PPROM, or IUGR in clinician-initiated delivery), and the uniform increase may be due the broad definition of sepsis in our study, which included early and late onset sepsis."

29. Page 21: the most likely misclassification would be the complicated cases of PPROM classified as iatrogenic deliveries, I doubt that this can be considered non-differential. Elaborate why you think that errors and omissions can result in underestimation of temporal trends.
RESPONSE: Even if complicated PPROM case were delivered by caesarean section, this would still be considered PPROM preterm birth rather than a clinician-initiated delivery because any record with PPROM indicator (see BERD data abstraction form) was primarily considered as PPROM preterm birth, regardless of the final mode of delivery. There is no evidence to suggest that PPROM would be diagnosed and classified differently by calendar year (e.g., in 2004 vs. 2012) as there has been no change in PPROM diagnostic criteria or data collection practices. Random errors and omissions increase year-to-year variability in the rates of preterm birth subtypes, which in turn increases noise-to-signal ratio that can obscure the temporal trend. Please, see also responses to comment #18.

Reviewer 2

Specific comments

1. As this is an international medical journal, the global readership will be interested in implications beyond the need for further research. What should the public health department, hospital management and clinicians be considering on the basis of this findings plus those of related studies? This should be included in the abstract and discussion.

RESPONSE: The most interesting findings in this study were (1) that the clinician-initiated preterm birth rate increased in every gestational age category, (2) that neonatal mortality decreased among infants born at 32-33 weeks following clinician-initiated delivery and increased among infants born at 34-36 weeks (late preterm) following clinician-initiated delivery, and (3) that composite neonatal mortality/morbidity increased in particular at late preterm regardless of preterm birth subtype. While increased clinician-initiated delivery appears to be associated with better outcomes for moderately preterm infants, the late preterm infants comprise the largest group of preterm infants. The implications for the public health department, hospital management and clinicians is to recognize the increased adverse outcomes and increased burden in terms the temporal changes in the proportion of vulnerable infants and appropriately distribute resources to order to accommodate the need. The rise in neonatal morbidity among late preterm infants was driven mainly by the increase in neonatal sepsis. This have implications for antibiotic use in the NICU as well as changes in antibiotic resistance and antimicrobial use and associated adverse neonatal outcomes (Ting, 2016 JAMA Peds).

We added the following to the manuscript:

Abstract: Page 2: “Timing of obstetric interventions is associated with infant health outcomes at preterm.”

Discussion: Page 19: “Unified diagnostic criteria and antimicrobial policies are needed to further examine and address this issue.”

2. The readership may not know about Washington State (e.g., vs Washington D.C.). The introduction should include a description of the state including population size, how it fares in preterm births and mortality compared with other states (e.g., where does it rank in the US). Likewise, the international readership would probably be interested how the US and the state measure up to other high-income (and maybe high-middle income) countries. The discussion would then benefit from putting the findings into broader context (neighboring states, US as a whole and global).

RESPONSE: The United States has one of the highest rates of preterm birth among industrialized countries. According to the March of Dimes Premature Birth Report Card, Washington State is one of the few states with an “A” rating and the lowest rates of preterm birth among the other states (https://www.marchofdimes.org/mission/prematurity-reportcard.aspx). The neonatal mortality rate in Washington State in 2004 was 3.3 per 1 000 live births and decreased to 3.0 per 1 000 live births in 2013. Which is lower than the overall neonatal mortality rate in the United States which decreased from 4.5 per 1 000 live births in 2004 to 4.0 per 1 000 live births in 2013 (https://www.marchofdimes.org/peristats/ViewSubtopic.aspx?reg=53&top=6&stop=107&lev=1&obj=1&cmp=99&slev=4&sty=2004&eny=2013&chy=).

We added a description of Washington State in the Discussion where we also discuss implications of national/international ranking with respect to preterm birth and neonatal mortality.

Page 21: “Washington State has one of the lowest preterm birth rates in the USA, and lowest infant mortality rates;(46) however, the ranking is very much dependent on the ethnicity, age, and socioeconomic status composition of the obstetric population.(46) We adjusted for number of these indices thus our results are relevant to other states in the U.S.A. and high-income countries in general.”

3. The readership will probably confuse iatrogenic and unnecessary procedures leading to preterm birth. In common usage in my areas of work, the two are almost synonymous. Obviously the critical importance is that unnecessary iatrogenic procedures that result in high risk of adverse outcomes are an immediate target for intense monitoring and enforcement of standards. However, the intro

describes it as "iatrogenic (medically-indicated)". This immediately prompts several questions: Are all iatrogenic preterm births truly medically indicated? How can the reviewers be sure? What safeguards are in place? (e.g., independent validation of records.) How did Washington State get around false reporting? (i.e., it being written as indicated when in fact it was not.) In the past two years, I have personal contacts with two persons who had unnecessary cesarean, one with a failed induction where the lie was require surgical intervention (admittedly not from Washington State). Bottom line, more description is needed that iatrogenic procedures were necessary. Otherwise, the proportions necessary or unnecessary should be included and reanalysis of proportion of adverse outcomes that could be prevented if unnecessary iatrogenic preterm births eliminated as well as the major implications of the unnecessary ones in abstract, discussion and conclusions. If these are not clearly available, this is a major limitation and should be stated as such. (This goes well beyond line 31-32 in the paragraph with limitations).
RESPONSE: The term “iatrogenic preterm birth” has been commonly used in the literature in contrast to “spontaneous preterm birth” to describe any non-spontaneous preterm birth resulting from labor induction or caesarean delivery (Morken, 2007 Paed Perinatal Epidemiol; Lisonkova, 2011 BMC Pregnancy Childbirth; Wong, 2011 Clin Perinatol; Joseph, 2013 Paed Perinatal Epidemiol; Fritz, Acta Obstet Gynecol Scand 2018). The term “medically-indicated preterm birth” is also used in the literature to refer to infants born iatrogenically due to maternal or fetal indications (Goldenberg, 2008 Lancet). However, the term “clinician-initiated preterm delivery” has become increasingly more common (Ananth, 2018 JAMA Pediatrics). The term “clinician-initiated preterm delivery” is the more appropriate term in context of our study, even though we believe that clinician-initiated delivery before 34 weeks and a vast majority of those between 34-36 weeks are truly medically indicated.

To avoid confusion of the readership, the term “iatrogenic” has been replaced with “clinician-initiated” in the manuscript.

4. "The completeness and accuracy of these databases was monitored by the Washington State Department of Health with annual assessments and consistency checks.17-19" The reviewer believes more description of the validation mechanisms. Has Washington State solved the age-old problem of inaccurate completion of death certificates and other reporting forms some of which were used in the study? This includes intentional and unintentional inaccurate reporting. If truly validated, then the accuracy (not the completeness) should be stated in the manuscript.
RESPONSE: Data used in this study were obtained from Linked Birth-CHARS files from the Washington State Department of Health. The Washington State Department of Health provides annual reports on the completeness of the Comprehensive Hospital Abstract Reporting System (CHARS) data, and these reports include the number of values missing or out-of-range. (http://www.doh.wa.gov/DataandStatisticalReports/HealthcareinWashington/HospitalandPatientData/HospitalDischargeDataCHARS/CHARSReports). The frequency of diagnostic and procedure codes is monitored in yearly reports, providing an insight into possible over-diagnosis of certain conditions or erroneous miscoding. Data correction is made for records that are flagged with inconsistent or out of range entries. The manual for data collection states the following: “Incorrect records remain in the production system pending hospital review and correction. Records with errors are displayed online. Saving a corrected record re-edits the record and any remaining errors are displayed. The process continues until all errors are corrected.”
(http://www.doh.wa.gov/Portals/1/Documents/5300/CHARSManual-UB04-5010.pdf). For consistency in reporting, Washington State hospitals have a standardized form to collect information for birth certificates (https://www.doh.wa.gov/Portals/1/Documents/Pubs/422-020-WashingtonStateBirthFilingForm.pdf). Birth certificate data are abstracted from the standardized form to the birth certificate form by trained data abstractors, while hospitalization data (CHARS) are abstracted for hospital billing and administrative purposes in the discharge abstract form by trained data abstractors. Validity studies show that completeness of information increases with data linkage

including both sources of information (birth certificates and hospitalization data), and data accuracy (sensitivity) also increases in linked files when two data sources report the same information. For example, in their validation study, Lydon-Rochelle et al. showed that data in Washington state linked datasets (birth certificates linked to hospital discharge data) had the highest true positive fraction (TPF) as compared with gold –standard of medical chart review by physician experts. For example, TPF for gestational diabetes was 93.3%, and occurrence of placenta previa had TPF 79.1%, when missing values were considered as 'condition not present' (Lydon-Rochelle, AJOG 2005). Another study of Lydon-Rochelle et al. used a sample of Washington State Linked Birth-CHARS files and showed that, for example, diagnostic code of major puerperal infection in hospital discharge data had positive predictive value 81.6% and negative predictive value 98.5% as compared with a gold-standard – a thorough chart review. A supporting sentence has been added to the Methods. Further, there is no reason nor evidence to suspect any intentional inaccurate reporting of health-related data. There will always be a trade-off between the inaccuracies of data in large datasets (i.e., population-based data) versus relatively accurate data in smaller hospital-based datasets (e.g., chart-abstracted data). We believe that both approaches add to the general knowledge and results should be reported in scientific journals.
The following text was added to the Methods:
Page 6: "The completeness and accuracy of these databases was monitored by the Washington State Department of Health with annual assessments and consistency checks.(16–18) Records flagged with inconsistent or out-of-range entries were addressed systematically through hospital review and correction. The frequency of diagnostic and procedure codes was monitored in annual reports.(18) Previous validation studies of the linked dataset showed that the positive and negative predictive values (PPV and NPV) for delivery characteristics were above 80% and 98%, respectively; (19,20) for example, labor induction had PPV 89.0% and NPV 94.5%. (20)"

5. The non-statistical specialist readership (and reviewer) may not understand why three periods of three years were included instead of an annual rate. This should be spelled out. If this is a standard, it should be explained and referenced
RESPONSE: The rates by period for the three periods were initially reported for descriptive purposes. These tables have been revised to include only the rates of the two most extreme periods (2004-2006 and 2011-2013) and the middle period (2007-2010) has been removed from the tables to avoid confusion. This is used to show a contrast between the earliest period and the most recent period. We chose to not use an annual rate for the preterm birth trends because the annual preterm birth rates is already shown more efficiently in the figures. We changed the tables to address the comments of both reviewers and be consistent in the reporting format.

6. The STROBE Statement is a very useful tool for both authors and reviewers. The reviewer recommends in addition to stating the place that may answer the statement, to put in quotes the actual statement.
RESPONSE: This is an important point about the usefulness of the STROBE Statement as a tool for both authors and reviewers. In addition to the page numbers where the statement is answered, the quotation of the actual statement has been added to the STROBE Statement tool for this manuscript and other manuscripts for submission in the future.

VERSION 2 – REVIEW

REVIEWER	Lorthe Elsa INSERM U1153, EPOPé team, France
REVIEW RETURNED	24-Sep-2018

GENERAL COMMENTS	I would like to thank the authors for the efforts put into considering my comments. The modified version of the article is satisfying. I only have very minor comments. Page 7: Why were severe congenital anomalies included in the analysis? For instance, most fetuses with anencephaly will be delivered as TOPs or will die shortly after birth. Page 10: “then” should be changed to “than”. Table 1 and others: P-value should be labelled as (period 2 vs period 1). Page 20: I still do not understand “iatrogenic termination of pregnancy”. A TOP necessarily results from a medical intervention, so is necessarily iatrogenic, right? I would remove the word “iatrogenic” to avoid readers’ misinterpretation.

REVIEWER	Howard Sobel Regional Coordinator for Reproductive, Maternal, Newborn, Child and Adolescent Health World Health Organization, Western Pacific Regional Office
REVIEW RETURNED	26-Sep-2018

GENERAL COMMENTS	No additional comments.

VERSION 2 – AUTHOR RESPONSE

Reviewer 1

1. I would like to thank the authors for the efforts put into considering my comments. The modified version of the article is satisfying. I only have very minor comments.
RESPONSE: Thank you for the time you took to review and provide insightful comments on our manuscript. We thank the reviewer for the additional thoughtful comments.

2. Page 7: Why were severe congenital anomalies included in the analysis? For instance, most fetuses with anencephaly will be delivered as TOPs or will die shortly after birth.
RESPONSE: We examined crude and adjusted trends in neonatal mortality, both of which included all live births. Adjustment for congenital anomalies aimed to examine if temporal changes in their birth prevalence would change our results, while keeping the study population (all live births as a denominator) consistent. Detailed analyses on the effects of congenital anomalies of temporal trends in neonatal mortality would require more details on individual anomalous conditions. We clarified the adjustment for congenital anomalies in the manuscript:
Page 7: “Information on congenital anomalies was obtained from BERD and included any of the following conditions observed within the first 24 hours after birth: anencephaly, meningomyelocele or spina bifida, cyanotic congenital heart disease, congenital diaphragmatic hernia, omphalocele, gastroschisis, limb reduction, cleft lip, cleft palate, Down syndrome, chromosomal disorders, and hypospadias. We adjusted for temporal trends in these conditions as a potential risk factor for adverse outcomes.”

3. Page 10: “then” should be changed to “than”
RESPONSE: The sentence was corrected as follows:
Page 10: “...more pregnancies occurred from assisted conception in 2011-2013 than in 2004-2006.”

4. Table 1 and others: P-values should be labelled as (period 2 vs period 1).
RESPONSE: This was corrected in Table 1 and Appendix Table 2.

5. Page 20: I still do not understand “iatrogenic termination of pregnancy”. A TOP necessarily results from a medical intervention, so is necessarily iatrogenic, right? I would remove the word “iatrogenic” to avoid readers’ misinterpretation.
RESPONSE: The word “iatrogenic” was removed to avoid readers’ misinterpretation, and the sentences were revised as follows:
Page 20: “Second, information on termination of pregnancy was not available; thus, we could not account for these temporal changes, However, the vast majority of pregnancy terminations are likely to occur prior to 24 week’s gestation; terminations beyond 23 weeks would be included as stillbirths in this study.”

Reviewer 2

1. No additional comments.
RESPONSE: Thank you for the time you took to review and provide insightful comments on our manuscript.