PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Maternal over-the-counter analgesics use during pregnancy and adverse perinatal outcomes: Cohort study of 151,141 singleton pregnancies
AUTHORS	Zafeiri, Aikaterini; Raja, Edwin Amalraj; Mitchell, Rod; Hay, David; Bhattacharya, Sohinee; Fowler, Paul

VERSION 1 – REVIEW

REVIEWER	Deborah Randall University of Sydney, Clinical and Population Perinatal Health Research
REVIEW RETURNED	04-Mar-2021

GENERAL COMMENTS	The study uses a large registry of pregnancies and births over a 30 year period in Aberdeen, UK, to examine the impact of over-the-counter analgesic use in pregnancy and adverse pregnancy outcomes. The authors point to the large number of births as a strength of this study, but I have concerns about whether this databank has enough information about the exposure and possible confounding by indication to accurately assess effect sizes and causality. I commend the authors on a complex piece of research work with the multiple outcomes and exposures, but I am concerned that the large size of the sample is giving precise but biased estimates, due to several issues that I mention below. 1. The authors state that OTC analgesic use was asked about in the first antenatal clinic visit. Seeing as most of the pregnancies had their first visit in the first trimester, the use must be predominantly first-trimester use, is that correct? If a woman has her first clinic visit in the third trimester, is the analgesia use question just for the first trimester, or does it cover the whole pregnancy? This needs to be clarified throughout the paper. Has this OTC analgesia question been validated? 2. If the adverse pregnancy outcomes are indeed caused by the OTC analgesics, and OTC analgesic use has increased so dramatically during the study period, we would expect to see an increase in these outcomes particularly in the 2005-2015 time period. What are the trends in these adverse outcomes, after accounting for changing demographics, pregnancy risk factors and also changes in obstetric practice (e.g. move to earlier iatrogenic births)? If there have not been unexplained increases in these adverse factors at the time when the OTC analgesia usage increased, then I would suggest that some of the effects reported here are due to unmeasured confounding and/or bias. With the exposure in this study a pregnancy-average exposure, or a

	trimester-averaged exposure, rather than an exposure that accounts for dose, duration and timing of exposure, I would have expected that the associations would be biased towards the null, which makes me wonder at some of the reported ORs, and whether they can be a true, unbiased estimate of the effect. 3. What happened in 2005-2015, to lead to such a dramatic increase in those taking at least one analgesic? Was this a gradual increase between 2005-2015 or was it due to diclofenac becoming available over the counter? I found it difficult to determine in the figure due to the log scale axis. Did the authors test an interaction between time period and exposure on outcomes, or do a stratified analysis? The authors looked at diclofenac use and outcomes for this last time period, but I would like to see crude and adjusted outcomes for the 'any analgesia' stratified by time period. 4. Those women who have taken at least one form of analgesia are much more likely to have a planned birth (Elective CS or Induction) and far less likely to have a SVD. Is this purely to do with the time frame, and the increasing proportion of iatrogenic births over time? This "practice" change towards earlier, planned births is happening in many places in the world, and is not fully accounted for by changing pregnancy risk factors, but may have a big impact on some of the outcomes like preterm birth, birthweight and Apgar scores. Due to the association between analgesic use and time period, the changing obstetric intervention practice at the same time could be a potential source of bias, and may not be fully accounted for by including the broad time periods in the regression analysis, particularly if there is an interaction. If the association between labour onset/mode of birth and OTC analgesia use is not related to the time period, then it would appear that the women with any OTC analgesia are a much higher risk group, as they are being treated very differently at birth. 5. This data should be analysed as a time-to-event analysis with time-varying covariates, such as a Cox proportional hazards model, otherwise all women are assumed to have the same exposure time, which is not the case, particularly for preterm birth/stillbirth outcomes. For example, some of the "baseline" characteristics are time-varying, but these are all treated as fixed. Characteristics like maternal hypertension and antepartum haemorrhage become more likely the longer the pregnancy continues. And this will affect the outcome rates such as stillbirth and preterm birth for women with these characteristics. This could be introducing bias in the adjusted analyses. 6. There are some adjusted ORs that are greater than the crude ORs (e.g. stillbirth, neonatal death, Apgar <7 at 1 min, Apgar <7 at 5min). The authors should investigate these further, and determine which variables are causing these changes in the OR away from the null, and whether there are interactions here that should be investigated. 7. I thank the authors for including their DAG. This helps to examine assumptions. I have some comments on the assumptions: - What is the hypothesised reason that parity is causally related to analgesia during pregnancy? - What is the hypothesised reason for maternal diet to impact on parity? - The authors draw a line from analgesia during labour to analgesia

	during pregnancy, but this would mean the future is impacting the past. - What about other factors that can influence outcomes such as chronic diseases? - I think that maternal pain, existing medical conditions, and other medications, are unmeasured confounders that could have a big impact on the association between OTC analgesia and the outcomes. This is a major source of potential confounding by indication. The authors acknowledge this in the Discussion, but I do not follow the argument on why this is not a major factor that affects the findings of this study. I would suggest, at minimum, that a quantitative bias analysis should be done to examine how much of an effect this could have, if this information was available. Also, I think there should be a DAG for each outcome, so that the different adjustment variable sets are justified. Some variables that are adjusted for in some outcome analyses are not in the current overall DAG. 8. The Conclusion in the Abstract states that the use of paracetamol in combination with other drugs confers the highest risk. Is this what is assessed in the main analysis? I thought it was “at least one”. The authors also mention having analgesia combinations as a key strength of the study, but I could not see combinations analysed, e.g. paracetamol and at least one other? 9. Table 1. When I add up the numbers taking paracetamol, ibuprofen/aspirin/naproxen, diclofenac, the numbers do not add up to the “at least one analgesic” column. E.g. in 2005-2015 I get n=22656, but the number taking at least one is 30998. Where do the remaining numbers in the “at least one analgesic” column come from? Minor comments: - May be worth putting acetaminophen in brackets when first introducing paracetamol. - Line 157 and Table 1: should be “obese >= 30” - Results: Lines 197-200 are very confusing. Please start with the proportion who reporting any otc analgesic use, and go from there. - Standardised birth weight score – authors say it is treated as a continuous variable, but then what groups are the OR and aOR comparing? Actually, seems as though it is not an OR, as there are negative values in the table. This is very confusing (and is stated as an OR in the text). - Table 3 – something wrong with the footnote, which is at the side of the table. - Discussion, line 309: “reaching 70.5% of women in the final decade of our study”. This is not the correct percentage - 70.5% of women with at least one OTC analgesia in the study period, are in this last time period (column %); 60.1% of women in 2005-2015 have taken at least one OTC analgesia (row %), up from 10.3% in 1985-1994. - Lines 310-312: I don’t follow what is meant by this sentence - “This means that our findings are applicable far beyond the percentage (between 14% and 38%) of pregnant women with underlying health deficits related to the adverse outcomes we report here.”

REVIEWER	May Ching Soh Oxford University Hospitals NHS Foundation Trust
REVIEW RETURNED	09-Mar-2021

GENERAL COMMENTS	I'd like to commend the authors on their efforts to address an area of study that is fraught with controversy by doing this study. It spans over a 30-year interval and prescribing practices may have evolved. However, the premise of this study hinges on self-reported analgesic use at the initial antenatal clinic appointment which is unverified (as this are over the counter drugs). This should be included in the title of the paper i.e. "Self-reported maternal analgesic use in pregnancy and adverse pregnancy outcomes.". I am concerned this cohort-based study where women were required to self-report analgesic use did not capture data on maternal comorbidities which may have necessitated analgesia use e.g. chronic rheumatic disease (in which potentially teratogenic disease-modifying therapies such as methotrexate and leflunomide would be co-prescribed) or other maternal medical conditions that could have significant impact on both obstetric and perinatal outcomes. Suggestions: 1. The exact duration of analgesic use is not captured in this study. Analgesic use could have been in very early pregnancy – which would affect organogenesis. And late use would affect fetal cardiac and renal function which was not reported in this study. 2. Aspirin was also included, and this could potentially be misleading. Low-dose aspirin is often used to modify the risk of placentally medicated adverse obstetric outcomes and therefore would be recommended for women who are at risk of these poor outcomes – which also happen to be the outcome variables collected in this study i.e. preterm delivery < 37 weeks, low birthweight < 2500g and stillbirth. I think it would only be fair if aspirin were removed from the analysis unless authors were able to demonstrate that this is not low dose aspirin used. (Therefore, it is unsurprising that the aspirin / naproxen and ibuprofen group demonstrated significant associations with poor placentally mediated outcomes of premature deliver, stillbirth. Prematurity is associated with admissions to neonatal / special care units. Would be more useful to give us median gestation delivered in this group. 3. This study spans a 30-year interval and the authors have rightly pointed out that analgesic prescription has evolved over time. It would be useful to divide the data into 5-7 yearly intervals (or in the date ranges used in Table 1) to see if the pattern of malformations has also evolved over time. 4. There are no data on maternal comorbidities other than smoking and BMI at baseline. Additional medical comorbidities like pre-existing renal disease, diabetes, autoimmune diseases are associated with poorer perinatal outcomes. As this is a registry-based study in a single centre, these data should be readily available and included in the analysis – in part of the multivariate analysis when determining adverse perinatal outcomes. 5. There are significant gaps in the data on which between 2005 – 2015 there were no data on paracetamol, aspirin, naproxen and ibuprofen use, whereas in those intervals data on diclofenac was available. Therefore, these data should not be comparable. I would suggest splitting this study into two – diclofenac vs. none and perhaps paracetamol vs. none. Numbers of women on aspirin, ibuprofen and naproxen are too small to be of interest. 6. I am particularly concerned with the authors' statement: "Our study demonstrates an association of maternal over-the-counter analgesic consumption during pregnancy with adverse neonatal offspring outcomes esp. when maternal comorbidities have not been included in this study." Statements such as these are unhelpful esp. with the general public who may shun the use of any analgesia in

	pregnancy. 7. Tables are too complicated most of the p-values have little meaning in some areas- please remove. a. Table 1 use of analgesia in labour, baby's presentation and delivery and sex of the baby are of little relevance to this paper and can be removed b. Maternal comorbidities such as pre-existing hypertension, renal disease, concurrent diseases like diabetes (type) autoimmune rheumatic diseases and inflammatory bowel disease, need to be included. c. Antenatal problems like maternal gestational diabetes needs to be included with hypertensive disorders 8. Giving (adjusted) odds ratios when the numbers affected are so small have very little meaning. The use of numbers needed to harm are more useful as it would help put into context some of these numbers for the reader.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Dr. Deborah Randall, University of Sydney Comments to the Author:

The study uses a large registry of pregnancies and births over a 30 year period in Aberdeen, UK, to examine the impact of over-the-counter analgesic use in pregnancy and adverse pregnancy outcomes. The authors point to the large number of births as a strength of this study, but I have concerns about whether this databank has enough information about the exposure and possible confounding by indication to accurately assess effect sizes and causality. I commend the authors on a complex piece of research work with the multiple outcomes and exposures, but I am concerned that the large size of the sample is giving precise but biased estimates, due to several issues that I mention below.

1. The authors state that OTC analgesic use was asked about in the first antenatal clinic visit. Seeing as most of the pregnancies had their first visit in the first trimester, the use must be predominantly first-trimester use, is that correct? If a woman has her first clinic visit in the third trimester, is the analgesia use question just for the first trimester, or does it cover the whole pregnancy? This needs to be clarified throughout the paper. Has this OTC analgesia question been validated?

The Discussion is amended to clarify this good point. OTC analgesics have not been biochemically validated and this is further clarified in the Methods.

Line 128: "Analgesic consumption was not further assessed analytically."

Lines 361-364: "Most women had their first antenatal clinic visit during the 1st pregnancy trimester, which might imply our results were affected by primarily 1st trimester exposure, although analgesic use in first trimester is most likely replicated in the rest of pregnancy."

2. If the adverse pregnancy outcomes are indeed caused by the OTC analgesics, and OTC analgesic use has increased so dramatically during the study period, we would expect to see an increase in these outcomes particularly in the 2005-2015 time period. What are the trends in these adverse outcomes, after accounting for changing demographics, pregnancy risk factors and also changes in obstetric practice (e.g. move to earlier iatrogenic births)? If there have not been unexplained increases in these adverse factors at the time when the OTC analgesia usage increased, then I would suggest that some of the effects reported here are due to unmeasured confounding and/or bias.

With the exposure in this study a pregnancy-average exposure, or a trimester-averaged exposure, rather than an exposure that accounts for dose, duration and timing of exposure, I would have expected that the associations would be biased towards the null, which makes me wonder at some of the reported ORs, and whether they can be a true, unbiased estimate of the effect.

65% of the dramatic increase in prevalence of use between 2005 and 2015 was due to diclofenac becoming available over-the-counter, and which is, therefore, only analysed for this period. All outcome analyses are adjusted for year of delivery. In a separate analysis, ORs for 2005-2015 were calculated to cross-check for potential temporal effects in all analyses. However, temporal bias could be excluded since the results were comparable for significant outcomes.

3. What happened in 2005-2015, to lead to such a dramatic increase in those taking at least one analgesic? Was this a gradual increase between 2005-2015 or was it due to diclofenac becoming available over the counter? I found it difficult to determine in the figure due to the log scale axis. Did the authors test an interaction between time period and exposure on outcomes, or do a stratified analysis? The authors looked at diclofenac use and outcomes for this last time period, but I would like to see crude and adjusted outcomes for the 'any analgesia' stratified by time period.

Page 15, lines 306-308: "Diclofenac use increased steeply from 2005 (Figure 2A), which reflects the change in Scottish legislation, leading to diclofenac becoming available over the counter, without prescription, in that year."

This phrase already included in the manuscript explains the first part of the comment. We consider the log scale plot to be clearer but are happy to leave the choice between log or linear plots to the editor. We have attached the non-logged version of the figure in this revised version of the manuscript.

4. Those women who have taken at least one form of analgesia are much more likely to have a planned birth (Elective CS or Induction) and far less likely to have a SVD. Is this purely to do with the time frame, and the increasing proportion of iatrogenic births over time? This "practice" change towards earlier, planned births is happening in many places in the world, and is not fully accounted for by changing pregnancy risk factors, but may have a big impact on some of the outcomes like preterm birth, birthweight and Apgar scores. Due to the association between analgesic use and time period, the changing obstetric intervention practice at the same time could be a potential source of bias, and may not be fully accounted for by including the broad time periods in the regression analysis, particularly if there is an interaction. If the association between labour onset/mode of birth and OTC analgesia use is not related to the time period, then it would appear that the women with any OTC analgesia are a much higher risk group, as they are being treated very differently at birth.

Analysis for APGAR scores is already adjusted for year of delivery, maternal age at delivery, SIMD, first gestational booking and type of delivery (see Tables 2 and 3). Preterm birth is adjusted for year of delivery, maternal age at delivery, SIMD, first gestational booking, maternal hypertensive disorders, maternal antepartum haemorrhage (see Tables 2 and 3). Birthweight is adjusted for year of delivery, maternal age at delivery, SIMD, first gestational booking, gestation at delivery (see Tables 2 and 3). This means we have accounted for temporal trends. Furthermore, the change in planned births as suggested by the reviewer would affect both exposed and non-exposed groups in our cohort.

5. This data should be analysed as a time-to-event analysis with time-varying covariates, such as a Cox proportional hazards model, otherwise all women are assumed to have the same exposure time, which is not the case, particularly for preterm birth/stillbirth outcomes. For example, some of the "baseline" characteristics are time-varying, but these are all treated as fixed. Characteristics like maternal hypertension and antepartum haemorrhage become more likely the longer the pregnancy continues. And this will affect the outcome rates such as stillbirth and preterm birth for women with these characteristics. This could be introducing bias in the adjusted analyses.

Data were collected from case notes and entered into the registry. Not all baseline variables in the database have reference to the time during pregnancy. Therefore, a time-to-event analysis and treatment of some covariates as time-varying was not performed as pregnancy only lasts a maximum of 40 weeks so time variance would be miniscule. This is one of the limitations of the study and we have mentioned it in our discussion section:

Lines 356-359: “Our data were based on medical notes; however, over-the-counter consumption is self-reported, and details on the timing, duration, dosage, product type (single-ingredient vs combination) and administration type were not available in the database.”.

6. There are some adjusted ORs that are greater than the crude ORs (e.g. stillbirth, neonatal death, Apgar <7 at 1 min, Apgar <7 at 5min). The authors should investigate these further, and determine which variables are causing these changes in the OR away from the null, and whether there are interactions here that should be investigated.

Adjusted ORs were indeed higher for stillbirth, neonatal death, admission to neonatal unit, APGAR score at 1 min and APGAR score at 5 min in comparison to the crude ORs. The multivariate models used for the analysis of these outcomes were repeated, adding the covariates used for adjustment one by one in order to determine which one had the effect on increasing the OR. It was the same variables for all outcomes that were increasing the value, and these were (1) year of delivery and (2) first gestational booking. This relationship is predictable since analgesic consumption increased dramatically over the duration of the study.

7. I thank the authors for including their DAG. This helps to examine assumptions. I have some comments on the assumptions:

- What is the hypothesised reason that parity is causally related to analgesia during pregnancy?

Parity is associated with an increased likelihood of analgesic use. This is an indicative reference for reviewer’s information: Harris, GM.E., Wood, M., Eberhard-Gran, M. et al. Patterns and predictors of analgesic use in pregnancy: a longitudinal drug utilization study with special focus on women with migraine. BMC Pregnancy Childbirth 17, 224 (2017). https://doi.org/10.1186/s12884-017-1399-0

- What is the hypothesised reason for maternal diet to impact on parity?

Nutrition, diet and obesity have a known effect on female fertility. We have provided an indicative list of references for the reviewer below:

Broughton DE, Moley KH. Obesity and female infertility: potential mediators of obesity's impact. Fertil Steril. 2017 Apr;107(4):840-847. doi: 10.1016/j.fertnstert.2017.01.017. Epub 2017 Mar 11. PMID: 28292619.

Hohos NM, Skaznik-Wikiel ME. High-Fat Diet and Female Fertility. Endocrinology. 2017 Aug 1;158(8):2407-2419. doi: 10.1210/en.2017-00371. PMID: 28586412; PMCID: PMC6283234.
Meldrum DR. Introduction: Obesity and reproduction. Fertil Steril. 2017 Apr;107(4):831-832. doi: 10.1016/j.fertnstert.2017.02.110. PMID: 28366410.

Silvestris E, Lovero D, Palmirotta R. Nutrition and Female Fertility: An Interdependent Correlation. Front Endocrinol (Lausanne). 2019 Jun 7;10:346. doi: 10.3389/fendo.2019.00346. PMID: 31231310; PMCID: PMC6568019.

- The authors draw a line from analgesia during labour to analgesia during pregnancy, but this would mean the future is impacting the past.

We thank the reviewer for spotting this error in the figure, which is now corrected.

- What about other factors that can influence outcomes such as chronic diseases?

These factors are included in “other medical condition” variable, which is unmeasured in our cohort.

- I think that maternal pain, existing medical conditions, and other medications, are unmeasured confounders that could have a big impact on the association between OTC analgesia and the outcomes. This is a major source of potential confounding by indication. The authors acknowledge this in the Discussion, but I do not follow the argument on why this is not a major factor that affects the findings of this study. I would suggest, at minimum, that a quantitative bias analysis should be done to examine how much of an effect this could have, if this information was available.

Previous maternal medical conditions and use of other medications in combination are unmeasured variables in our study. Most chronic conditions with pain as one of the symptoms are usually treated

with prescription of analgesics, including stronger compounds not included in our study. The list of unmeasured variables in an epidemiological cohort is always larger than what can be measured, for example we could argue that these pregnant women are also exposed to BPA, breathing air pollution, eating fish and many other exposures that can also affect our outcomes. In this study we are specifically looking to calculate whether in utero exposure of the fetus to over-the-counter analgesic compounds is associated with an increased incidence of the studied outcomes, adjusting for confounders to the best of our abilities and database availability.

Also, I think there should be a DAG for each outcome, so that the different adjustment variable sets are justified. Some variables that are adjusted for in some outcome analyses are not in the current overall DAG.

During our attempt to address this comment, generating and including 12 DAG tables (one for each outcome as suggested), resulted in 12 almost identical tables to be added in supplementary material. We believe this adds very little towards reader's understanding of our analysis since the tables are so similar to our overall table. Therefore, we have not included 12 separate DAG tables.

8. The Conclusion in the Abstract states that the use of paracetamol in combination with other drugs confers the highest risk. Is this what is assessed in the main analysis? I thought it was "at least one". The authors also mention having analgesia combinations as a key strength of the study, but I could not see combinations analysed, e.g. paracetamol and at least one other?

As shown by the numbers in our tables, and discussed in text, paracetamol was used by the majority of cases and so it inevitably has the biggest influence on the results from "at least one" group (lines 119-122). This can be also seen from the comparable results from the analysis of the paracetamol only group. The "at least one" group includes cases with consumption of at least one analgesic out of the 5 tested in this study, meaning it can include cases where 1 drug was used, 2 drugs, 3 drugs etc. up to all 5 together. Therefore, potential analgesic drug combinations in these groups can be simply calculated as 2^5=32 possibilities. However, since we are not counting the choice of none of the five analgesics, we subtract 1 to get 31 possible combinations of analgesics. We also analysed and reported a combination of drugs separately for our group "Ibuprofen/Aspirin/Naproxen" for further clarity.

9. Table 1. When I add up the numbers taking paracetamol, ibuprofen/aspirin/naproxen, diclofenac, the numbers do not add up to the "at least one analgesic" column. E.g. in 2005-2015 I get n=22656, but the number taking at least one is 30998. Where do the remaining numbers in the "at least one analgesic" column come from?

Our "at least one" means one pregnancy can have more than 2 drugs so in this group the pregnancy with at least one is only reported once. Therefore, if a woman has taken paracetamol and ibuprofen is 1 case in "at least one" group, 1 case in paracetamol group and 1 case in ibuprofen group.

Minor comments:

- May be worth putting acetaminophen in brackets when first introducing paracetamol.

Acetaminophen is already included in our list of keywords. We have now also amended the text as suggested (line 71).

- Line 157 and Table 1: should be "obese >= 30"

This has been corrected.

- Results: Lines 197-200 are very confusing. Please start with the proportion who reporting any otc analgesic use, and go from there.

We have modified the text as suggested.

- Standardised birth weight score – authors say it is treated as a continuous variable, but then what groups are the OR and aOR comparing? Actually, seems as though it is not an OR, as there are negative values in the table. This is very confusing (and is stated as an OR in the text).

Linear regression was used to explore relationship between analgesic use and Standardized birth weight. We have, therefore, reported the difference (p value) in Standardized birth weight score between offspring of women who used analgesic and women who did not use analgesic, after controlling for other covariates, and its 95% CI. We have added a footnote on tables 2 and 3 and edited the text to further increase clarity.

- Table 3 – something wrong with the footnote, which is at the side of the table.

Unfortunately, the footnote is in the correct location in our version, so we cannot correct.

- Discussion, line 309: "reaching 70.5% of women in the final decade of our study". This is not the correct percentage - 70.5% of women with at least one OTC analgesia in the study period, are in this last time period (column %); 60.1% of women in 2005-2015 have taken at least one OTC analgesia (row %), up from 10.3% in 1985-1994.

We have amended the text as suggested:

Lines 301-302: "reaching 70.5% of women using those compounds in the final decade of our study."

- Lines 310-312: I don't follow what is meant by this sentence - "This means that our findings are applicable far beyond the percentage (between 14% and 38%) of pregnant women with underlying health deficits related to the adverse outcomes we report here."

We would indicate that in the preceding sentence we stated: "There was a dramatic increase in the frequency of over-the-counter (non-prescription) analgesic use in pregnancies between 1985 and 2015, reaching 70.5% of women using those compounds in the final decade of our study." We would expect that in our population somewhere between 14% and 38% of the women would have had an underlying condition, as indicated in lines 303-304. Given that 70.5% of women the last decade of our study were using OTC analgesia, our findings are applicable to at least another 31.5 of women above the maximum of 39% with underlying conditions. The only conclusion, therefore, is that our findings are applicable to pregnant women with and without underlying conditions.

Reviewer: 2
Dr. May Ching Soh, Oxford University Hospitals NHS Foundation Trust Comments to the Author:
I'd like to commend the authors on their efforts to address an area of study that is fraught with controversy by doing this study. It spans over a 30-year interval and prescribing practices may have evolved. However, the premise of this study hinges on self-reported analgesic use at the initial antenatal clinic appointment which is unverified (as this are over the counter drugs). This should be included in the title of the paper i.e. "Self-reported maternal analgesic use in pregnancy and adverse pregnancy outcomes.".

I am concerned this cohort-based study where women were required to self-report analgesic use did not capture data on maternal comorbidities which may have necessitated analgesia use e.g. chronic rheumatic disease (in which potentially teratogenic disease-modifying therapies such as methotrexate and leflunomide would be co-prescribed) or other maternal medical conditions that could have significant impact on both obstetric and perinatal outcomes.

Suggestions:
1. The exact duration of analgesic use is not captured in this study. Analgesic use could have been in very early pregnancy – which would affect organogenesis. And late use would affect fetal cardiac and renal function which was not reported in this study.

The Discussion has been amended to include mention of this limitation.

Lines 356-xxx: “Our data were based on medical notes; however, over-the-counter consumption is self-reported, and details on the timing, duration, dosage, product type (single-ingredient vs combination) and administration type were not available in the database.”

2. Aspirin was also included, and this could potentially be misleading. Low-dose aspirin is often used to modify the risk of placentally medicated adverse obstetric outcomes and therefore would be recommended for women who are at risk of these poor outcomes – which also happen to be the outcome variables collected in this study i.e. preterm delivery < 37 weeks, low birthweight < 2500g and stillbirth. I think it would only be fair if aspirin were removed from the analysis unless authors were able to demonstrate that this is not low dose aspirin used. (Therefore, it is unsurprising that the aspirin / naproxen and ibuprofen group demonstrated significant associations with poor placentally mediated outcomes of premature deliver, stillbirth. Prematurity is associated with admissions to neonatal / special care units. Would be more useful to give us median gestation delivered in this group.

The Discussion has been amended to include mention of this limitation.

Lines 359-361: “In addition, the group of pregnancies with aspirin consumption might include use of low-dose aspirin which is recommended to help reduce risk of some pregnancy complications and outcomes related to placental function.”

3. This study spans a 30-year interval and the authors have rightly pointed out that analgesic prescription has evolved over time. It would be useful to divide the data into 5-7 yearly intervals (or in the date ranges used in Table 1) to see if the pattern of malformations has also evolved over time.

Sub-analyses for all our analyses have been previously performed for years 2005-2015 separately. The results were comparable for all significant outcomes, which further confirms that no temporal effect is present.

4. There are no data on maternal comorbidities other than smoking and BMI at baseline. Additional medical comorbidities like pre-existing renal disease, diabetes, autoimmune diseases are associated with poorer perinatal outcomes. As this is a registry-based study in a single centre, these data should be readily available and included in the analysis – in part of the multivariate analysis when determining adverse perinatal outcomes.

Apart from maternal smoking status and BMI, other maternal comorbidities included in our tables are SIMD (Scottish Index of Multiple Deprivation) and hypertensive disorders which in turn include gestational hypertension and preeclampsia. We agree with the reviewer that the additional comorbidities mentioned in their comment can be associated with perinatal outcomes, however these variables need to also be associated with our exposure in order to affect outcomes as a biasing factor. In addition, these conditions are proportionately at a low frequency and it is unlikely they would make a significant difference in our analysis.

5. There are significant gaps in the data on which between 2005 – 2015 there were no data on paracetamol, aspirin, naproxen and ibuprofen use, whereas in those intervals data on diclofenac was available. Therefore, these data should not be comparable. I would suggest splitting this study into two – diclofenac vs. none and perhaps paracetamol vs. none. Numbers of women on aspirin, ibuprofen and naproxen are too small to be of interest.

We suggest that the reviewer may have misinterpreted the table. The only data gap is for diclofenac which was not authorised as an OTC in Scotland until 2005. As shown in Table 1, there are data for 2005-2015 for all drugs. The fact that diclofenac was only used between 2005-2015 was the reason for the sub-group analysis shown in Table 3 which is only for that period of time. The years have been added in the title of Table 3 so it is also there and not just in text.

6. I am particularly concerned with the authors’ statement: “Our study demonstrates an association of maternal over-the-counter analgesic consumption during pregnancy with adverse neonatal offspring outcomes esp. when maternal comorbidities have not been included in this study.” Statements such as these are unhelpful esp. with the general public who may shun the use of any analgesia in pregnancy.

This is what the data shows, ours is not the only study showing associations between analgesic exposure and adverse neonatal outcomes , but, to our knowledge is currently one of the largest, with data on the largest range of analgesics over a long study time. We do not conclude and neither do we state that women should stop using OTC analgesics.

7. Tables are too complicated most of the p-values have little meaning in some areas- please remove.
a. Table 1 use of analgesia in labour, baby's presentation and delivery and sex of the baby are of little relevance to this paper and can be removed

Baby's presentation at delivery is an indicator of potential complications during labour, and therefore relevant to be mentioned as background data. Given considerable sex-specific differences that have been reported, the sex of the baby is also relevant to our study since we are analysing sex-specific outcomes as well. In addition, removing our results for these two out of fourteen baseline characteristics would not have much impact on simplifying the table.

b. Maternal comorbidities such as pre-existing hypertension, renal disease, concurrent diseases like diabetes (type) autoimmune rheumatic diseases and inflammatory bowel disease, need to be included.

See above answer to Reviewer 1 on this.

c. Antenatal problems like maternal gestational diabetes needs to be included with hypertensive disorders

We are not certain we understand the meaning for this addition.

8. Giving (adjusted) odds ratios when the numbers affected are so small have very little meaning. The use of numbers needed to harm are more useful as it would help put into context some of these numbers for the reader.

We thank the reviewer for this suggestion. Numbers needed to harm have now been calculated for all analyses and are included in the manuscript as Supplementary Table 1 and Supplementary Table 2. The text is also amended accordingly:

Lines 184-185: "Numbers needed to harm (NNH) were also calculated for each outcome and are provided in Supplementary Tables 1 and 2."

Lines 288-291: "As most of the outcomes studied were relatively rare the numbers needed to harm were mostly more than 100. Preterm birth, low birthweight and admission to the neonatal unit were exceptions with NNH ranging from 15 to 38. (Tables S1 and S3)."

Lines 324-328: ". Most numbers needed to harm for our outcomes (Tables S1 and S2) ranged between 1000 and 100, apart from preterm birth, low birth weight and baby admission to neonatal unit, which were 27, 38 and 15 respectively for our main analysis, further strengthening observed associations."

There also needs to be a clear statement in the Discussion as to how this means that our findings ARE of concern.

VERSION 2 – REVIEW

REVIEWER	Deborah Randall University of Sydney, Clinical and Population Perinatal Health Research
REVIEW RETURNED	08-Oct-2021

GENERAL COMMENTS	I would like to thank the authors for responding to my initial comments, and reiterate that I believe this to be an important area of

	study and I commend the authors on tackling this topic. However, many of my concerns remain. While the study has a large sample, and information on specific types of analgesia use, there are some major issues in the design with the main ones being: 1. Lack of information on specific exposure time and duration and dose. 2. Lack of information on indication for use. In the recent consensus statement in Nature Reviews Endocrinology (Bauer et al 2021) on paracetamol use in pregnancy, they call for better epidemiological research in this area, but point out that what is needed are studies that are designed to: “reduce confounding by indication for use”, “control for genetic factors”, “accurately capture exposure and outcome”, and “examine timing, dosage and duration of exposure both prenatally and postnatally”. The design of this analysis does not include these specific elements that are called for. I think the fact that this study cannot address these design criteria must be more explicitly acknowledged in the introduction and discussion. At this stage, the authors appear to suggest that they do not believe confounding by indication could be explaining these results, but I do not agree. Also for further elaboration in the discussion section: - One of the adverse outcomes that the consensus paper considers potentially likely to be linked with paracetamol use is cryptorchidism. However, this is one outcome where, after adjustment, there is no association with paracetamol. Why do the authors think that this specific adverse outcome that has a possible plausible causal pathway through endocrine disruption is not significantly associated with paracetamol use in their study, when other, more general adverse outcomes are? - Please take more space in the results to compare findings with the Rebordosa et al (2009) study, which was also a large population-based cohort with a sample size of almost 100,000, and had more detailed information on trimester of paracetamol usage and also collected information on indication for use. They did not find any association with paracetamol use and risk of stillbirths, low birth weight or small size for gestational age. Increased risk of preterm birth was only for those with pre-eclampsia and paracetamol use in the third trimester. What do the authors believe are the reasons for the different results? Further modifications that I would suggest to the authors include: - I would suggest removing the “at least one medication” analysis, and focus on the separate medications, or specific combinations, e.g. paracetamol on its own or with diclofenac. The way the medications impact on the pregnancy and the fetus may be quite different, so putting them together does not really help with assess the main causal questions. - I would like to see sensitivity analyses in supplementary materials that stratify by the three broad 10-year groups (still adjusting for a year variable within these broad year groups of course). - I will suggest again that a quantitative bias analysis would help to understand what the potential impact of unmeasured confounding could be on the results and conclusions, particularly the missing information on indications for use. - I thank the authors for attempting a DAG for every outcome, but I do believe this is important. Each outcome will have different confounders and mediators, and it is important that these have been

	considered by the authors. Some comments on the current overall DAG – where does gestational age fit in, as this is adjusted for in some of the analyses. Are other indications for use captured under 'other/previous medical conditions'? This should have a direct link with outcomes, as well as through other medications. I think type of delivery should also have a direct link with outcomes. Pre-eclampsia/hypertension is also a tricky one, as this seems to have been associated with paracetamol use in other studies, but it could be reverse causation. Additional comments: - Please detail further in the methods how year of delivery, maternal age at delivery, SIMD and maternal first antenatal visit were adjusted for in the analyses. As categorical variables (which categories, if so)? As continuous variables? If not done already, I would suggest adjusting for maternal age and year of delivery as continuous variables if possible, accounting for possible non-linear relationships as needed, e.g. using splines. - Table 1: Analgesia during labour – Is this including only those who had analgesia during vaginal birth? The % is lower than the % who had a caesarean section. Shouldn't this include all those who had a caesarean section too? - Table 1: Baby presentation row titles are reversed, I believe. Further discussion of points from the first review: Initial comment: 9. Table 1. When I add up the numbers taking paracetamol, ibuprofen/aspirin/naproxen, diclofenac, the numbers do not add up to the "at least one analgesic" column. E.g. in 2005-2015 I get n=22656, but the number taking at least one is 30998. Where do the remaining numbers in the "at least one analgesic" column come from? Author response: Our "at least one" means one pregnancy can have more than 2 drugs so in this group the pregnancy with at least one is only reported once. Therefore, if a woman has taken paracetamol and ibuprofen is 1 case in "at least one" group, 1 case in paracetamol group and 1 case in ibuprofen group. Additional comment: I did not initially realise that the paracetamol column was 'only' paracetamol, although I see now that this is in the table title, and in the methods. I would suggest putting this in the column title too, to make it really clear. Is the diclofenac analysis diclofenac only as well? Please include this in the table title. Initial comment: - Discussion, line 309: "reaching 70.5% of women in the final decade of our study". This is not the correct percentage - 70.5% of women with at least one OTC analgesia in the study period, are in this last time period (column %); 60.1% of women in 2005-2015 have taken at least one OTC analgesia (row %), up from 10.3% in 1985-1994. Author response: We have amended the text as suggested: Lines 301-302: "reaching 70.5% of women using those compounds in the final decade of our study." Additional comment: I would suggest that it is the usage in the current population that is more pertinent, and I would suggest reporting the 60.1% figure instead. And I'm afraid I still do not understand what the following sentence is getting at. Is it about indications for use? Please explain further, and

	if talking about prevalence in the overall pregnant population, please compare against the 60% of women in that time period who used at least one analgesic, not the 70% of all analgesic users who are in that time period.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 1

Dr. Deborah Randall, University of Sydney Comments to the Author:

I would like to thank the authors for responding to my initial comments, and reiterate that I believe this to be an important area of study and I commend the authors on tackling this topic. However, many of my concerns remain. While the study has a large sample, and information on specific types of analgesia use, there are some major issues in the design with the main ones being:

1. Lack of information on specific exposure time and duration and dose.

2. Lack of information on indication for use.

In the recent consensus statement in Nature Reviews Endocrinology (Bauer et al 2021) on paracetamol use in pregnancy, they call for better epidemiological research in this area, but point out that what is needed are studies that are designed to: “reduce confounding by indication for use”, “control for genetic factors”, “accurately capture exposure and outcome”, and “examine timing, dosage and duration of exposure both prenatally and postnatally”. The design of this analysis does not include these specific elements that are called for.

I think the fact that this study cannot address these design criteria must be more explicitly acknowledged in the introduction and discussion. At this stage, the authors appear to suggest that they do not believe confounding by indication could be explaining these results, but I do not agree.

We are grateful to Dr Randall for taking the time to review and critique our manuscript. In line with her comments the text in introduction and discussion have been amended to acknowledge the limitations in our study.

Lines 94-97: “Here we address many limitations -however, not all22 - of previous studies by analysing one of the largest cohorts, widest range of health data and, pregnancy use of five over-the-counter analgesics consumed in combination or separately.“

Lines 351-353: “A potential concern was that women were probably using analgesics to treat some inherent medical condition which in turn could have been the mediating factor for adverse outcomes. Data on indication for use were not available in the database.”

Lines 360-363: “Our data were based on medical notes; however, over-the-counter consumption is self-reported, and details on the timing, duration, dosage, product type (single-ingredient vs combination) and administration type were not available in the database.”

Lines 365-367: “Genetic factors potentially relating to the emergence of offspring health outcomes was an unmeasured variable in our analysis.”

Also for further elaboration in the discussion section:

- One of the adverse outcomes that the consensus paper considers potentially likely to be linked with paracetamol use is cryptorchidism. However, this is one outcome where, after adjustment, there is no association with paracetamol. Why do the authors think that this specific adverse outcome that has a possible plausible causal pathway through endocrine disruption is not significantly associated with paracetamol use in their study, when other, more general adverse outcomes are?

We consider that this point has been addressed in the discussion of the limitations of our study.

Lines 374-377: “Compared to our cohort size, there were, overall, very few cases of cryptorchidism, neural tube defects, amniotic band defects, hypospadias and gastroschisis, potentially underpowered statistical analyses to detect a difference for these outcomes.”

- Please take more space in the results to compare findings with the Rebordosa et al (2009) study, which was also a large population-based cohort with a sample size of almost 100,000, and had more detailed information on trimester of paracetamol usage and also collected information on indication for use. They did not find any association with paracetamol use and risk of stillbirths, low birth weight or small size for gestational age. Increased risk of preterm birth was only for those with pre-eclampsia and paracetamol use in the third trimester. What do the authors believe are the reasons for the different results?

The text in discussion has been amended to include the comparison with the findings of this paper in more detail:

Lines 399-408: “Another study with a large sample size (98,190 pregnancies) and a 7 year study time from Rebordosa and colleagues29, also reported an increased risk of preterm birth following paracetamol use during pregnancy, which was increased in mothers with pre-eclampsia. Our results showed a significant association of the adjusted ORs following adjustment for maternal hypertensive disorders. In addition, they did not find a significant association with stillbirth, or low birth weight as we report here. This disagreement could be due to dataset differences including the information about use in each pregnancy trimester, but also methodological differences such as the use of questionnaires versus medical notes or adjustment for different confounders.”

Further modifications that I would suggest to the authors include:

- I would suggest removing the “at least one medication” analysis, and focus on the separate medications, or specific combinations, e.g. paracetamol on its own or with diclofenac. The way the medications impact on the pregnancy and the fetus may be quite different, so putting them together does not really help with assess the main causal questions.

We thank the reviewer for this comment. We would like to point out that our “at least one analgesic” analysis provides additional information on the results presented in this study. It also better encapsulates real-life consumption where use might be alternating between two or more different compounds at the same time. This analysis includes all potential combinations of use of these 5 analgesics.

- I would like to see sensitivity analyses in supplementary materials that stratify by the three broad 10-year groups (still adjusting for a year variable within these broad year groups of course).

We believe that performing and including these analyses stratified by the three 10-year groups while also adjusting for year of delivery will introduce some over-adjustment in our results. Years have been categorised into these groups to account for changes in analgesics consumption over the years along with obstetrics care advancements, while we are already adjusting for year of delivery in all our analyses.

- I will suggest again that a quantitative bias analysis would help to understand what the potential impact of unmeasured confounding could be on the results and conclusions, particularly the missing information on indications for use.

The text in discussion has been amended to include mention of this limitation.

Lines 367-368: “This study does not include a quantitative bias analysis to identify potential distort of results presented here.”

- I thank the authors for attempting a DAG for every outcome, but I do believe this is important. Each outcome will have different confounders and mediators, and it is important that these have been considered by the authors. Some comments on the current overall DAG – where does gestational age fit in, as this is adjusted for in some of the analyses. Are other indications for use captured under ‘other/previous medical conditions’? This should have a direct link with outcomes, as well as through other medications. I think type of delivery should also have a direct link with outcomes. Pre-eclampsia/hypertension is also a tricky one, as this seems to have been associated with paracetamol use in other studies, but it could be reverse causation.

We have included the DAG diagrams for each outcome separately, as requested, in supplementary material along with our 1st version categorising all outcomes together as “perinatal offspring outcomes”.

Text in methods has also been amended to include mention of the supplementary figures:

Lines 171-172: “The socio-demographic characteristics that were likely to confound our exposure-to-outcome path were identified using directed acyclic graphs (DAG) (Supplementary figures S1-11)25.”

Additional comments:

- Please detail further in the methods how year of delivery, maternal age at delivery, SIMD and maternal first antenatal visit were adjusted for in the analyses. As categorical variables (which categories, if so)? As continuous variables? If not done already, I would suggest adjusting for maternal age and year of delivery as continuous variables if possible, accounting for possible non-linear relationships as needed, e.g. using splines.

Already included in methods text, lines 143-156: “The baseline characteristics compared between exposed and unexposed pregnancies were (reference category first): year of delivery (1985-1994, 1995-2004, 2005-2015), maternal age at delivery (20-25, <20, 26-35, >35 years), previous pregnancy (no; yes), maternal body mass index (BMI) (normal weight 18.5-24.9 kg/m2, underweight <18.5 kg/m2, overweight 25-29.9 kg/m2, obese >30 kg/m2), maternal first antenatal visit (1st, 2nd, 3rd trimester), maternal smoking status (non-smoker, smoker, ex-smoker), Scottish Index of Multiple Deprivation (SIMD) decile (1-6, 7-10, decreasing deprivation with increasing score), maternal hypertensive disorders (no disorder, gestational hypertension, preeclampsia, eclampsia), maternal antepartum haemorrhage (no haemorrhage, abruption, placental previa), type of labour (spontaneous, elective caesarean section, induced), type of delivery (spontaneous vaginal delivery, instrumental, caesarean section), analgesia during labour (no; yes), baby presentation at delivery (occiput anterior, occiput posterior), baby sex (female; male).”

- Table 1: Analgesia during labour – Is this including only those who had analgesia during vaginal birth? The % is lower than the % who had a caesarean section. Shouldn’t this include all those who had a caesarean section too?

The “analgesia during labour” variable includes all pregnancies regardless of delivery type.

- Table 1: Baby presentation row titles are reversed, I believe.

We are grateful to the reviewer for spotting this. We have double-checked Table 1 based on this comment and we confirm that this was a formatting error. Row titles have now been corrected.

Further discussion of points from the first review:

Initial comment:

9. Table 1. When I add up the numbers taking paracetamol, ibuprofen/aspirin/naproxen, diclofenac, the numbers do not add up to the “at least one analgesic” column. E.g. in 2005-2015 I get n=22656, but the number taking at least one is 30998. Where do the remaining numbers in the “at least one analgesic” column come from?

Author response:

Our “at least one” means one pregnancy can have more than 2 drugs so in this group the pregnancy with at least one is only reported once. Therefore, if a woman has taken paracetamol and ibuprofen is 1 case in “at least one” group, 1 case in paracetamol group and 1 case in ibuprofen group.

Additional comment:

I did not initially realise that the paracetamol column was ‘only’ paracetamol, although I see now that this is in the table title, and in the methods. I would suggest putting this in the column title too, to make it really clear. Is the diclofenac analysis diclofenac only as well? Please include this in the table title.

This information is mentioned as fully as possible within the methods, results, and discussion text, and in table titles:

Lines 122-125: “Then, three sub-group analyses against the control group were performed using only paracetamol, only diclofenac, or at least one analgesic from aspirin/naproxen/ibuprofen as exposure groups, excluding pregnancies exposed to multiple analgesics at the same time (Figure 1).”

Lines 252-254: “In the sub-group analysis considering only paracetamol consumption during pregnancy as our exposure group, most of the associations reported in the main analysis remained significant with the same direction of significance (Table 2).”

Lines 298-300: “Consumption of paracetamol alone was further associated with higher odds for APGAR score <7 at one and five minutes both in crude and adjusted analyses.”

Lines 313-314: “Furthermore, and surprisingly, exposure to diclofenac only was associated with significant decrease in the incidence of stillbirth.”

Table 2 title: “Regression analysis of offspring outcomes between control (no analgesic) and groups exposed to at least one analgesic, only paracetamol, and at least one from ibuprofen, aspirin, naproxen”

Table 3 title: “Sub-group regression analysis between control pregnancies and exposed to diclofenac”

As helpfully suggested by the reviewer, we have now also included the word “only” for the relevant columns referring to these analyses in tables 1, 2 and 3.

Initial comment:

- Discussion, line 309: “reaching 70.5% of women in the final decade of our study”. This is not the correct percentage - 70.5% of women with at least one OTC analgesia in the study period, are in this last time period (column %); 60.1% of women in 2005-2015 have taken at least one OTC analgesia (row %), up from 10.3% in 1985-1994.

Author response: We have amended the text as suggested:

Lines 301-302: “reaching 70.5% of women using those compounds in the final decade of our study.”

Additional comment:

I would suggest that it is the usage in the current population that is more pertinent, and I would suggest reporting the 60.1% figure instead.

We agree with the reviewer that reporting the 60.1% figure instead of the 70.5% is clearer.

And I’m afraid I still do not understand what the following sentence is getting at. Is it about indications for use? Please explain further, and if talking about prevalence in the overall pregnant population, please compare against the 60% of women in that time period who used at least one analgesic, not the 70% of all analgesic users who are in that time period.

In order to reduce potential misunderstanding, we have amended the text as follows.

Lines 300-304: “There was a dramatic increase in the frequency of over-the-counter (non-prescription) analgesic use in pregnancies between 1985 and 2015, starting from only 10.3% of women using 1 or more of the compounds between 1985 and 1994, climbing to 60.1% of women in the final decade of our study.”