PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Antidepressant use during pregnancy and the risk of major congenital malformations in a cohort of depressed pregnant women: An updated analysis of the Quebec Pregnancy Cohort.
AUTHORS	Berard, Anick; Zhao, Jin-Ping; Sheehy, Odile

VERSION 1 - REVIEW

REVIEWER	Angela Lupattelli University of Oslo, Norway
REVIEW RETURNED	31-Aug-2016

| GENERAL COMMENTS | This is an important study that adds considerably to the analysis of the effects of first trimester antidepressant use on the risk of congenital malformations. A critique of other studies has been that there is a problem in comparing women with depression and antidepressant use with women without depression or associated psychiatric diagnoses, because the increased risks of malformations associated with antidepressant use might reflect either differences due to the psychiatric diagnosis or due to unmeasured confounders such as increased rates of smoking among women with psychiatric diagnoses.

Another critique was the inclusion of minor as well as major malformations as there may be more of a detection bias with minor malformations if a woman is known to be taking antidepressants or other prescription medications. Secondly, analyses of malformation rates for a shorter period after birth will miss those that are not diagnosed until later.

This study addresses all of these concerns. However, I had two main suggestions related to the analysis. One is to control the analyses for use of benzodiazepine and benzodiazepine-like drugs. if this is not a pre-planned analysis it could be reported as an exploratory secondary analysis. My rationale is that Oberlander et al. found an increased rate of malformations with SSRIs + benzodiazepine use but not SSRIs alone. Additionally, benzodiazepine exposure is fairly frequent in pregnancy and in a cohort with depression/anxiety diagnoses.

Secondly, given the large numbers and range of diagnoses included to construct this sub-cohort of pregnant women with a psychiatric diagnosis related to mood or anxiety, it would be helpful to include an additional adjustment for the severity level of the diagnosis, if only in broad categories, assuming a standardized scale or some guidance exists for such division. This again could be done as a secondary exploratory analysis. It would address the potential concern that within this large set of diagnoses, those with |
|---|---|

antidepressant exposure in pregnancy may be at a more severe illness level. Additionally, it would address the higher known smoking rates among patients with severe psychiatric illness. The lack of differences in psychiatric visits, except for TCA users and also the likelihood that some first trimester use is inadvertent ongoing use rather than a sign of severe illness, suggests that this may not be the case. However, it is worth carrying out an exploratory analysis to address this, given the large differences in severity of included diagnoses.

Related to this - is it possible to separate out ED visits, hospitalizations and psychiatric hospitalizations in the year pre-pregnancy, for the cohort as a whole? Given the frequency of hospitalizations or ED visits combined, this might provide more of a picture of how similar or different these groups were.

Some more minor points:
- please avoid the acronym MCM. As major congenital malformations are a key concept and this is not a general readership commonly used acronym, it should be spelled out in full for better readability;
- the wording 'were increasing the risk' or 'is increasing the risk' ' was statistically significantly increasing' should be changed, in part to improve the English ('were increasing' would be changed to 'increased') and in part for greater care to avoid claiming causation when an association has been observed; this is the case in the abstract and at various places in the text; this is the main reason I've noted a problem with the English; the English is fine otherwise.

- given the references to the effects of amitriptyline and other TCAs on serotonin in the text and abstract, a reference is needed to support the statement in the text on effects on serotonin uptake;
- page 7, avoid the acronym "1DG"
- page 8, I was surprised at the decision to exclude pregnancies with multiple antidepressant exposures, rather than analyzing them as a separate category. It would be worth noting that total numbers were small, and therefore individual cells (e.g. specific types of combinations) were too small for analysis if this is the case.
- p11: 'the number of medication used other than antidepressants" should be "the number of medications..."
- Table 4: Given the large number of analyses presented in this table, it is likely to raise a concern that some associations would be expected by chance. One option in addressing this would be to add a column with a much tighter confidence interval than 95%. A second might be to distinguish between a small number of hypothesized associations based on the existing research on antidepressants and various forms of major malformations (such as paroxetine and cardiac malformations) and other exploratory analyses. This would then require adjustment for the secondary analyses. The table would need to be re-organized in this case, with hypothesized associations listed first, then additional organ systems below. This suggestion has implications for the related results reporting as well. Additionally, are not ventricular/atrial septal defects a subset of cardiac malformations? If the latter category includes overall cardiac malformations, this should be clear in the table design and a footnote. I did not understand the note that "results are provided when defects were identified for the antidepressant type." Does this mean that comparisons were only carried out if at least one user of a specific antidepressant had a child with the specific type of malformation?

- STROBE: "in pregnancy" is not the study design. The title does include the statement that this is a cohort study.
- Figure 1: The text in this Figure could be simplified by not repeating information in the reasons for exclusions and the inverse of the reason in the inclusion column. This might also allow for fewer boxes in the inclusion column, without information loss. Text could also be shortened per box or reason for exclusion.
- Figure 2: given that diagnoses in the year pre-pregnancy were measured during this entire period to construct the cohort and also that inadvertent use in early pregnancy and intended ongoing use during pregnancy among women with depression/anxiety diagnoses probably make up the large majority of first trimester exposures, rather than new depression/anxiety diagnoses during pregnancy, it would be useful to know whether there has been a growth in diagnoses of depression in the year pre-pregnancy (or in the year pre-pregnancy and during pregnancy, in combination, if desired). This would be the more relevant information than only in-pregnancy diagnoses, as listed on Figure 2.

Please check the scale and data on this figure. I was surprised to see that the number of diagnoses in pregnancy exceeded antidepressant exposure in pregnancy. It is not clear if this is first trimester exposure only or exposure throughout pregnancy from the legend. Also the number reported as exposed in the first trimester in Figure 1 (36,440) represent 1.25% of the cohort of 289,688 after a number of exclusions. In Figure 2, the annual listed exposure rate grows from ~ 0.2% to 0.42%. This is inconsistent.

The discussion mentions the concordance of the study results with other studies in administrative data that have examined rates of malformations among users of antidepressants, and specifically findings for paroxetine and cardiac malformations, citalopram and cranyosynostosis, and venlafaxine and respiratory defects. In addition to discussing the coherence of the current study with previous studies, it would be valuable to know how the effect size in the current study compared with those in studies that failed to restrict the comparison group to women with a mood or anxiety disorder diagnosis. If effect sizes are similar, this provides additional strength to the likelihood that there is little confounding by indication in relation to evidence of malformation rates with first trimester antidepressants. This finding would be especially important if it is possible to also control for broad categories of diagnoses, in order to add a severity component to the analysis, and for concurrent benzodiazepine use as recommended above.

As noted above, this is an important study that uses a large data set to evaluate malformation rates among pregnant women with antidepressant exposure in comparison with unexposed pregnant women with similar psychiatric diagnoses. It addresses an important question raised in this research domain concerning the influence of confounding by indication and unmeasured confounders on observed malformation rates, and adds considerably to the existing body of research evidence and our understanding of antidepressant effects.

REVIEWER	Angela Lupattelli
	University of Oslo, Norway
REVIEW RETURNED	03-Aug-2016

GENERAL COMMENTS	Thank you for giving me the opportunity to review this paper. This is a study within the Quebec pregnancy cohort exploring whether gestational exposure to antidepressants increases the risk of major malformations. Generally, this is an important research question with relevant clinical implications, especially in light of the discordant literature on the topic; however there are important drawbacks that impede a correct interpretation of the findings.
	There is paper published on the American Journal of Obstetrics and Gynecology (2015) presenting data on sertraline alone and using the same data sources as in the current paper, with the sole difference of the time period covered (data extraction for two extra years in the sertraline study, i.e. 1998-2010). Beyond the concern of redundant publication in relation to sertraline, these two papers lead to different conclusions regarding sertraline and this could be misleading for the readers and clinicians. Is it plausible that just two additional years in the analyzed data would produce such findings? I guess this could have something to do with sample size and study power, and therefore why not using data up to 2010 in the current paper if these additional data are available?
	One major concern I have is about lack of sensitivity and probabilistic bias analyses, multiple testing, no information about missing data and how these were handled, and lack of translation of relative measures into absolute terms. By reading the STROBE checklist, it seems like these aspects were not taken into account. When studying important topics such as teratogenicity of medications, we cannot simply rely on the main results without any substantiation from sensitivity analyses.
	Here below are more specific comments on the various sections of the paper.
	Introduction The authors are encouraged to provide some magnitude estimates of the associations identified in previous studies between antidepressants and major congenital anomalies. Since various meta-analyses on the topic are available, it would be good to present/cite their findings.
	When dealing with previous study findings, it would be appropriate to comment on the quality of these study, methodological strength, rather than positive/negative findings. This could in a way address, at least in some extent, why the current information in the literature is still controversial.
	Methodology Please specify whether the inclusion criterion in the study, i.e. pregnancies exposed to antidepressants a year before conception, refers to antidepressant-exposed pregnancies where such use was in relation to a diagnosis of depression or anxiety. This is an important point to clarify since antidepressants could be used for other indications. In this case, the included study population would be more heterogeneous.
	Please indicate in the main text or as footnote in the flow-chart what "known teratogens" refer to, i.e., what medications were considered as major teratogens.
	Definition of exposure: it seems like the authors defined as exposed

pregnancies with an antidepressant prescription filled in the period from conception to end of first trimester. However, this could lead to a misclassification of exposure if a woman redeeming a prescription before conception, still uses antidepressants in early pregnancy. How did the authors deal with this? It would be important to know whether there has been any validation study of the Prescription Insurance Database specifically looking at this issue, and in relation to antidepressants. If not, the authors are encouraged to explore exposure misclassification and the related potential/direction of bias.

It is not clear in the text whether the substance level analyses were also for monotherapy; ie where these pregnancies only exposed to a single antidepressant during first trimester, or did any switching occur?

I encourage the authors to run sensitivity analyses with antidepressant exposure based on two dispensing in first trimester. This approach would give us more confidence on the results. Women filling twice a prescription are probably more likely to have taken the medication.

Definition of Outcome:
How did you deal with the heart problems among premature babies? i.e., cases of patent ductus arteriosus (PDA)? PDA is common in premature children but resolves afterwards; so, this cannot be classified as a heart defect. You should have at least run sensitivity analyses restricted to children not born prematurely in order to explore this research question. It is not clear in the main text whether this point was addressed at all.

It is also unclear to me whether malformation diagnoses were based on a single diagnosis code in infant record. Why not doing sensitivity analyses where infants need to have at least two recordings of the malformation diagnosis? The prevalence of major malformations is; it should at least be addressed in the Discussion, and it would be relevant to know the prevalence when the diagnosis is based on at least two diagnoses on the infant record.

I would suggest several additional sensitivity analyses that would make us more confident about the findings, for example examining only first pregnancy for women participating more than once in the cohort, or restricting the analysis to term delivery.

The authors state that major malformations have been validated in the RAMQ and MedEcho against patient charts, however there is no information about PPV and NPV. The authors should provide the reader with this info if available and are encouraged to run sensitivity analyses to explore the possibility of outcome misclassification.

Statistical analysis:
The authors are encouraged to include a power or sample size calculation for their study.

The authors should consider expand a bit the rationale behind the selection of covariates used in the adjusted model. Adjustment for many variables has been done, which may lead to over-adjustment. Use of directed acyclic graphs to define the minimally sufficient set of confounders is encouraged here. It is strange that there was no adjustment for comedication use in first trimester (e.g., NSAIDs,

anxiolytics). It seems like the models were adjusted for "other prescribed drugs in the year prior to LMP", but what is the rationale for not including comedications in first trimester?

The authors lack information on fundamental variables, such as folic acid use, smoking, and alcohol intake. These are very important to account for, and probabilistic bias analysis for unmeasured factors should be performed in this regard. Later in the Discussion the author state that these characteristics are probably similar between antidepressant users and non-users, but several studies indicate the contrary.

It is also stated that "by design" the study controls for maternal depression, however having a diagnoses of depression one year conception does not give us information about the underlying severity of depression. I think this should be acknowledged in the Discussion section/study limitations.

Did the author have any information about the number of ultrasounds performed during pregnancy across the groups?

Results
Please address the main differences (sociodemograophics) between pregnancy excluded and included in the analysis.

It would be better to use the wording "depressed untreated/non-medicated" rather than "unexposed" in the Tables.

The authors should at least state the % of missing data on the explored covariates, and how missing data were handled. I could not find any information about it.

Discussion
The first paragraph of the Discussion should have a more realistic tone. Other previous studies have attempted to restrict the study cohort to women with depression. Also, it is not so beneficial to study all SSRIs as a group, since we know that individual substances can have diverse safety profiles in pregnancy in relation to congenital anomalies.

The authors should address more extensively the live-birth bias, and how this could have affected their estimates.

The authors acknowledge that several tests were run and thus some findings could be due to chance. However, how many findings in the current paper could be due to chance? Also, why nothing was done to address multiple testing is that was of concern?

The Discussion should be elaborated in light of the results of the necessary sensitivity and probabilistic bias analyses. Without these additional analyses, we cannot correctly interpret the study findings.

The authors are also encouraged to discuss their findings in absolute terms (absolute risk, NNH). Relative measures are not so informative, especially in the teratology field when clinicians (and also pregnant women) have to balance the risk of taking an antidepressant versus the risk of untreated depression for both mother and child health.

Strengths and Limitations – bullet point section

	The authors should list more limitations. First, they did not have information on essential confounders such as folic acid, smoking, alcohol, and did not adjust for co-medication in pregnancy. The live birth bias is another concern, and last but not least lack of information on depression severity.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Angela Lupattelli
Institution and Country: University of Oslo, Norway
Competing Interests: None declared

Thank you for giving me the opportunity to review this paper. This is a study within the Quebec pregnancy cohort exploring whether gestational exposure to antidepressants increases the risk of major malformations. Generally, this is an important research question with relevant clinical implications, especially in light of the discordant literature on the topic; however there are important drawbacks that impede a correct interpretation of the findings.

There is paper published on the American Journal of Obstetrics and Gynecology (2015) presenting data on sertraline alone and using the same data sources as in the current paper, with the sole difference of the time period covered (data extraction for two extra years in the sertraline study, i.e. 1998-2010). Beyond the concern of redundant publication in relation to sertraline, these two papers lead to different conclusions regarding sertraline and this could be misleading for the readers and clinicians. Is it plausible that just two additional years in the analyzed data would produce such findings? I guess this could have something to do with sample size and study power, and therefore why not using data up to 2010 in the current paper if these additional data are available?
Response: We thank the reviewer for pointing this out. A typographical error occurred in this manuscript. The same calendar years were used for this study and in the sertraline paper referred to by the reviewer (Bérard et al, 2015, AJOG). Hence, we have corrected the follow-up period from 1998-2008 to 1998-2010. The analyses between the two studies are different however given the greater number of exposure categories in the current study. We believe that the findings of this new submitted manuscript are more clinically applicable. Indeed, in this new manuscript, we have looked at specific antidepressant types instead of pooling all non-sertraline SSRI and non-SSRI antidepressants together, which is a clinical advantage as stated in the introduction. Hence, although sertraline users are the same, they are analyzed with different exposure categories.
We disagree however that we are reaching different conclusions. Indeed, if we look at the point estimates in our current manuscript and in Bérard et al. (AJOG 2015), the point estimates (overall and organ specific) are very similar, which add validity to our current analyses. Given the different analyses and exposure categories, we would not expect to have identical confidence intervals. The interpretation however is the same and in accordance with Bérard et al. (AJOG, 2015). We have added the following sentence to that effect in our discussion section on p.16, 4th paragraph: 'Reassuringly, findings on sertraline from Berard et al.10 have been replicated in this study although the confidence intervals do not reach statistical significance. The different inclusion criteria and exposure groups considered in the statistical analyses can partly explain this.'

One major concern I have is about lack of sensitivity and probabilistic bias analyses, multiple testing, no information about missing data and how these were handled, and lack of translation of relative measures into absolute terms. By reading the STROBE checklist, it seems like these aspects were not taken into account. When studying important topics such as teratogenicity of medications, we

cannot simply rely on the main results without any substantiation from sensitivity analyses.
Response:
We agree with the reviewer that when studying important topics such as teratogenicity of medications, we cannot simply rely on the main results without any substantiation from sensitivity analyses. As described throughout the manuscript, the Quebec Pregnancy Cohort is one of the most validated pregnancy cohorts in the World. Indeed, gestational age has been validated (Vilain et al., ref. 17), exposure status has been validated (Jobin-Gervais et al – ref 20), and major congenital malformation codes have been validated with patient charts (Blais et al., ref. 21). To counteract and adjust for the effect of potentially important covariates, the study cohort has been restricted to depressed pregnant women. This is a way to adjust for indication AND unmeasured confounders – Reviewer 2 agrees with us on this.

With regards to missing data: We disagree with the reviewer because we have mentioned this and taken into account the missing variables per design. This has been explained in the Method section and in the Discussion section on p. 17, last paragraph: 'Limitations include missing information on potentially important confounders such as smoking, folic acid intake, and alcohol intake. Although we could not adjust for these variables specifically, we adjusted for these per design by studying only depressed/anxious pregnant women. Indeed, all these variables are associated with depressive status16 regardless of antidepressant use. Given that it may take many months before obtaining an appointment to a psychiatrist in Quebec, data on psychiatrist visits were used as a proxy for severity of disease rather than presence of depression; diagnoses of depression with or without antidepressant use is a better suited definition for the presence of depression, which was used in this study. It was also reassuring that SSRI and SNRI users were similar to non-users with regards to the number of visits to psychiatrists in the year before pregnancy.' Hence, we could not provide information on these variables because we have none (unmeasured). We have taken this into account by studying a group of depressed/anxious women (restricted study group)– it is known that depressed pregnant women smoke, drink alcohol and take their folic acid in a similar way regardless of depressive status (ref 16 in the paper). We have gone further to adjust for severity status in order to limit residual confounding by depression severity. Again, reviewer 2 has praised this.

With regards to multiple testing: We agree that chance findings are always an issue in studies such as ours. All authors (us and others) mentioned in our discussion section have handled the potential for chance finding the same way as we did. Indeed, we had already acknowledged the possibility of chance finding on p. 18, 1st paragraph. For clarity, we have added the underlined text below: 'In addition, given the number of comparisons made in this study, we cannot rule out the fact that some of our findings could have occurred by chance alone. It is however reassuring that many of the statistically significant associations that we found have been reported by other authors before, decreasing the possibility of chance finding.'

We agree with the reviewer that we should consider findings in terms of absolute terms. We have therefore added the following section on p. 18-19, in the Conclusion: '… The baseline risk of major malformations is between 3-5% in the general population and individualized treatment remains warranted….' We do not want however to minimize the risk given the potential lack of efficacy as stated on the last sentence of the manuscript. Caution is warranted in this situation.

Here below are more specific comments on the various sections of the paper.

Introduction
The authors are encouraged to provide some magnitude estimates of the associations identified in previous studies between antidepressants and major congenital anomalies. Since various meta-analyses on the topic are available, it would be good to present/cite their findings. When dealing with previous study findings, it would be appropriate to comment on the quality of these study, methodological strength, rather than positive/negative findings. This could in a way address, at least in some extent, why the current information in the literature is still controversial.

Response: We thank the reviewer for this comment. Although we had mentioned this in the Introduction already, we have rephrased the second paragraph on p. 6 for clarity: 'Human studies concerning the adverse effects of antidepressant exposure during gestation on the developing fetus have showed increased risk of various congenital malformations such as cardiac, musculoskeletal, respiratory, cranyosynostosis, and cranyofacial,5-10 but variations between study results remain.11-15 Underlying maternal depression, unaccounted potential confounders, class effect compared to antidepressant type effect or lack of statistical power could potentially explain these.' Because of space restriction, we could not review all studies specifically in the Introduction but the most important published literature as well as meta-analyses (Bérard et al. (ref. 33), Wurst et al. (ref. 31)) findings are already discussed in the Discussion section. Given the potential limitations of such meta-analyses on the topic, we preferred discussing the original studies (see Discussion section – studies discussed and compared to our findings).

Methodology
Please specify whether the inclusion criterion in the study, i.e. pregnancies exposed to antidepressants a year before conception, refers to antidepressant-exposed pregnancies where such use was in relation to a diagnosis of depression or anxiety. This is an important point to clarify since antidepressants could be used for other indications. In this case, the included study population would be more heterogeneous.
Response: We agree with the reviewer and thank her for pointing this out. We have corrected the specific inclusion criterion as follow on p. 7-8: '…2) pregnancies with a diagnosis of depression and/or anxiety (Table S1, Supplemental files), and exposed to antidepressants in the 12 months before pregnancy'….

Please indicate in the main text or as footnote in the flow-chart what "known teratogens" refer to, i.e., what medications were considered as major teratogens.
Response: We have addressed this by referring to Briggs et al.18 and Kulaga et al.,19 in our manuscript. Like in all the other studies performed within the Quebec Pregnancy Cohort, we exclude pregnancies exposed to known teratogens/feto-toxic medications according to Kulaga et al. (ref. 19 – study performed with collaboration of Briggs) as well as Briggs et al. (a well known and reputable text; ref. 18). The list is too long to describe and thus we are referring to the list published in Kulaga et al. We could put it in a Supplemental file if requested by BMJ. We would need to get a copyright agreement from BJOG. Please confirm the request and we will ask for permission.
p. 8, 1st paragraph: 'We excluded pregnancies exposed to known teratogens during the 1st trimester of pregnancy according to Briggs et al.18 and Kulaga et al.,19 and pregnancies with newborn diagnoses of chromosomal abnormalities.'
We have however added one supplemental file (Table S6) describing the fetotoxic medications used by the 552 excluded pregnancies (excluded because of exposure to fetotoxic drugs (see Figure 1)). This has been also mentioned on p.10 of the Results section: '…a description of pregnancies excluded due to exposure to fetotoxic drugs is presented on Table S6 of the supplemental files.'

Definition of exposure: it seems like the authors defined as exposed pregnancies with an antidepressant prescription filled in the period from conception to end of first trimester. However, this could lead to a misclassification of exposure if a woman redeeming a prescription before conception, still uses antidepressants in early pregnancy. How did the authors deal with this? It would be important to know whether there has been any validation study of the Prescription Insurance Database specifically looking at this issue, and in relation to antidepressants. If not, the authors are encouraged to explore exposure misclassification and the related potential/direction of bias.
Response: We agree with the reviewer. This has already been mentioned in the Method section. Exposure has been validated ('Data on prescription fillings have been validated (overall exposures, and exposure to AD specifically)) and compared to maternal reports, which are more reliable than

data on medication prescribing in medical charts; the positive predictive value of prescription drug data in the cohort was found to be at least 87% (95%CI: 70%-100%) and the negative predictive value was at least 92% (95%CI: 86%-98%).20 The relevant exposure time-window was the first trimester confirmed by ultrasound. P. 10-11) – the underlined text has been added for clarity.
In addition, we have added the following underlined text on p. 10: 'We identified prescription fillings for any antidepressants (AD) dispensed to women in the study cohort from the Quebec Public Prescription Drug Insurance database, with the timing of exposure determined by the dispensed date and duration of prescription. The relevant exposure time-window was the first trimester (0-14 weeks of gestation) confirmed by ultrasound. Hence, pregnancies with prescriptions filled during the first trimester or prescriptions filled before pregnancy but with duration overlapping the first day of the last menstrual period were defined as exposed.'

It is not clear in the text whether the substance level analyses were also for monotherapy; ie where these pregnancies only exposed to a single antidepressant during first trimester, or did any switching occur?
Response: Only monotherapy. This was already mentioned in the Method section inclusion/exclusion criteria on p.8: 'Within this pre-defined cohort of depressed/anxious pregnancies, we further considered pregnancies that were exposed to only one type of antidepressants or non-exposed to antidepressants during the first trimester of pregnancy. This excluded pregnancies with multiple different antidepressant exposures during organogenesis, which are likely to be at increased risk of adverse pregnancy outcomes; this resulted in very few excluded pregnancies because the total number of combined AD uses or switches were small, and therefore individual cells (e.g. specific types of combinations) did not allow analyses.'

I encourage the authors to run sensitivity analyses with antidepressant exposure based on two dispensing in first trimester. This approach would give us more confidence on the results. Women filling twice a prescription are probably more likely to have taken the medication.
Response: Again, if exposure had not been validated, we would have agreed with the reviewer. However, we have taken good care to validate our exposure status before undertaking this study (Jobin-Gervais et al., ref. 20) and do not feel that it is necessary here.

Definition of Outcome:
How did you deal with the heart problems among premature babies? i.e., cases of patent ductus arteriosus (PDA)? PDA is common in premature children but resolves afterwards; so, this cannot be classified as a heart defect. You should have at least run sensitivity analyses restricted to children not born prematurely in order to explore this research question. It is not clear in the main text whether this point was addressed at all. It is also unclear to me whether malformation diagnoses were based on a single diagnosis code in infant record. Why not doing sensitivity analyses where infants need to have at least two recordings of the malformation diagnosis? The prevalence of major malformations is; it should at least be addressed in the Discussion, and it would be relevant to know the prevalence when the diagnosis is based on at least two diagnoses on the infant record.
Response: Very few premature births occurred – the mean gestational age in all exposure groups, including the non-exposed group was 38 weeks (see Table 1). As mentioned in the Method section, major malformations have been identified in the first year of life to make sure that diagnoses do not resorb spontaneously and to take into account delayed reporting:
p. 9: ..'Detection in the first year of life was accounted for to allow for late detection as well as negative confirmatory diagnoses.' Furthermore, as stated in the Method section and Discussion section (strengths), major congenital malformations have been validated (Blais et al., ref 21). We added the following sentence on p. 10, Method section for clarity: '…The positive predictive value of major congenital malformations diagnosed in the first year of life in the QPC have been found to be at least 80% and the negative predictive value 93%.21'

I would suggest several additional sensitivity analyses that would make us more confident about the findings, for example examining only first pregnancy for women participating more than once in the cohort.

Response: Only considering the first pregnancy would i) limit sample size (well known fact that it limits our ability to identify statistically significant findings), ii) would not reflect clinical reality – women with multiple pregnancies adjust their lifestyles according to the outcome of their previous pregnancy – this needs to be captured in the analyses given that risk factors between pregnancy differ – clinical need, iii) our analyses (GEE) take into account the different risk factors for each pregnancy as well as the 'mother clustering – underlying genetics', and iv) GEE is a well known and accepted method for rare outcomes such as major congenital malformations – it is used by other authors.

We would suggest that in order to be confident of the findings, one needs to compare study results to already published findings as well as animal data (biological plausibility) – we have done all of this in the Discussion section. One also needs to validate study variables (with validation studies of with study design/analyses) as much as possible – we have done this (see our response above), Furthermore, our findings on sertraline are very similar to what we had found in Berard et al. (2015), which is again reassuring. See also our answers to Reviewer 2.

Statistical analysis:

The authors are encouraged to include a power or sample size calculation for their study.

Response: In such a study, no sample size calculation is done because we use all subjects meeting inclusion/exclusion criteria within a pre-defined cohort with no possibility of recruiting other subjects. To answer the reviewer's comment however, we have provided post-hoc power calculations on p. 18 – Discussion section, limitation: '... Even if our estimates are based on high numbers of exposed cases, we cannot eliminate increased risk in analyses on specific malformations due to lower statistical power for some organ specific groups. Indeed, depending on the specific organ and antidepressant type analyzed, the post-hoc statistical power ranged from 28-86%...'

The authors should consider expand a bit the rationale behind the selection of covariates used in the adjusted model. Adjustment for many variables has been done, which may lead to over-adjustment. Use of directed acyclic graphs to define the minimally sufficient set of confounders is encouraged here. It is strange that there was no adjustment for comedication use in first trimester (e.g., NSAIDs, anxiolytics). It seems like the models were adjusted for "other prescribed drugs in the year prior to LMP", but what is the rationale for not including comedications in first trimester?

Response: All potential confounders considered were risk factors for major malformations or makers of disease severity. The last sentence of p. 10 has been revised as follows: 'Potential confounders were considered for all analyses if they were risk factors for congenital malformations or makers of disease severity:…'.

Co-medication during the first trimester was adjusted for (see p. 11, end of 1st paragraph). We have further presented co-medication use differently following Reviewer 2's comments below. We have separated benzodiazepines use from the other co-medications (Tables 1 and 2); ED, MD and hospitalizations are also presented more extensively in Tables 1 and 2 following Reviewer 2's comments. This did not change findings given that all these were already considered in the statistical analyses.

The authors lack information on fundamental variables, such as folic acid use, smoking, and alcohol intake. These are very important to account for, and probabilistic bias analysis for unmeasured factors should be performed in this regard. Later in the Discussion the author state that these characteristics are probably similar between antidepressant users and non-users, but several studies indicate the contrary.

Response: We would respectfully ask the reviewer to see our answer to this question in another one of her comment above. In addition, although it is true that antidepressant users are different from non-users in the general population, it is not the case in this study because all pregnant women were depressed. This is substantiated with a reference in our paper and agreed by Reviewer no. 2.

It is also stated that "by design" the study controls for maternal depression, however having a diagnoses of depression one year conception does not give us information about the underlying severity of depression. I think this should be acknowledged in the Discussion section/study limitations.
Response: Again, we would respectfully ask the reviewer to see our answer to this question in another one of her comment above. We have addressed severity of depression above.

Did the author have any information about the number of ultrasounds performed during pregnancy across the groups?
Response: Yes, we do have access to the number of ultrasounds. This is not a concern here because the number of ultrasounds during pregnancy (in the mother) and after pregnancy in the children are similar among depressed pregnant women and their children (Bar-Oz et al., ref 40 – this was taken from the Quebec Pregnancy Cohort).
We added the following sentences in the Discussion section on p. 18 for clarity: '… Given the general awareness on the risk of major malformations associated with antidepressant use during pregnancy, detection bias could be an issue. Nevertheless, the fact that only depressed women were studied here decreases this likelihood. Indeed, Bar-Oz et al.40 showed that depressed pregnant women and their children have similar number of ultrasounds during and after pregnancy.'

Results
Please address the main differences (sociodemograophics) between pregnancy excluded and included in the analysis.
Response: Our study was performed within a defined cohort of depressed pregnant women. We did not exclude any women BUT some pregnancies were not considered because they did not meet eligibility criteria (see Figure 1). Nevertheless, we have included a table in the supplemental files (Table S5) describing maternal age, place of living, and welfare status of the pregnancies that were not considered, stratified by the reason for exclusion as suggested by the Reviewer. In addition, as suggested by reviewer 1, a list of pregnancies excluded because of exposures to fetotoxic drugs is also presented in Table S6.
The following sentence was added on p. 12, 1st paragraph: 'Compared to pregnancies that were excluded, the study population was slightly younger and less likely to be on welfare (see Supplemental Table S5); a description of pregnancies excluded due to exposure to fetotoxic drugs is presented on Table S6 of the supplemental files.'
The following sentence was also added to the Discussion section (p.18-19, last sentence of Discussion section): 'Similarly, although the study population slightly differed from pregnancies not meeting eligibility criteria in terms of SES, this does not affect interval validity.'

It would be better to use the wording "depressed untreated/non-medicated" rather than "unexposed" in the Tables.
Response: Although we do not have data on other forms of treatments (other than antidepressants), pregnancies that are not using antidepressants might be treated with psychotherapy, light therapy, etc. Hence, they are not untreated. In addition, in order to compare our findings with other studies on the same topic, we believe that 'unexposed' will allow a better comparison – it is the same label as others use.

The authors should at least state the % of missing data on the explored covariates, and how missing data were handled. I could not find any information about it.

Response: We are working with a pre-defined study cohort and have complete data coverage unless stated otherwise (i.e. smoking, folic acid, alcohol use). Please refer to our previous answers.

Discussion
The first paragraph of the Discussion should have a more realistic tone. Other previous studies have attempted to restrict the study cohort to women with depression. Also, it is not so beneficial to study all SSRIs as a group, since we know that individual substances can have diverse safety profiles in pregnancy in relation to congenital anomalies.
Response: We have toned down the first part of the first sentence of the Discussion as: 'To our knowledge, our study is one of a few to specifically investigate the risk of major malformations associated with the use of antidepressant types during the first trimester of pregnancy in a cohort of depressed pregnant women; we are also the first to investigate the effect of antidepressants with serotonin reuptake inhibition as a group.' We have not changed the second part of the sentence because we are the first to consider serotonin reuptake drugs as a whole – we do not mean SSRI as a class but drugs with serotonin reuptake inhibition action.

The authors should address more extensively the live-birth bias, and how this could have affected their estimates.
Response: We have already addressed this in the Discussion section p. 18: '…Only liveborns were considered, as is the case in all studies like ours. However, it is known that antidepressants are increasing the risk of spontaneous abortions.38,39 Because spontaneous abortion is a determinant of severe malformations,23 our results are likely underestimates of the true risk….'

The authors acknowledge that several tests were run and thus some findings could be due to chance. However, how many findings in the current paper could be due to chance? Also, why nothing was done to address multiple testing is that was of concern?
Response: We agree that this is a potential issue. We are asking the reviewer to see our previous answers to her comment on this above.

The authors are also encouraged to discuss their findings in absolute terms (absolute risk). Relative measures are not so informative, especially in the teratology field when clinicians (and also pregnant women) have to balance the risk of taking an antidepressant versus the risk of untreated depression for both mother and child health.
Response: We have done this as suggested. We ask the reviewer to please see our response above.

Strengths and Limitations – bullet point section
The authors should list more limitations. First, they did not have information on essential confounders such as folic acid, smoking, alcohol, and did not adjust for co-medication in pregnancy. The live birth bias is another concern, and last but not least lack of information on depression severity.
Response: We have addressed all of these in the reviewer's comments above.

……………………………………
Reviewer: 2
Reviewer Name: Barbara Mintzes
Institution and Country: University of Sydney, Australia
Competing Interests: none declared

This is an important study that adds considerably to the analysis of the effects of first trimester antidepressant use on the risk of congenital malformations. A critique of other studies has been that there is a problem in comparing women with depression and antidepressant use with women without depression or associated psychiatric diagnoses, because the increased risks of malformations

associated with antidepressant use might reflect either differences due to the psychiatric diagnosis or due to unmeasured confounders such as increased rates of smoking among women with psychiatric diagnoses.

Response : We agree with Dr Mintzes that previous studies are limited by potential indication bias or unmeasured confounder bias as stated in our introduction. We therefore limited our cohort to depressed/anxious women to adjust for these two limitations by study design. This has been stated in the introduction already and as an answer to Reviewer 1 above.

Another critique was the inclusion of minor as well as major malformations as there may be more of a detection bias with minor malformations if a woman is known to be taking antidepressants or other prescription medications. Secondly, analyses of malformation rates for a shorter period after birth will miss those that are not diagnosed until later.

Response : Again, we agree with Dr Mintzes. These points have been stated already in our Method section (p. 10: 'Detection in the first year of life was accounted for to allow for late detection as well as negative confirmatory diagnoses.'; p.8: 'We further excluded pregnancies resulting in minor malformations alone in newborns. This was done because minor malformations are likely diagnosed selectively (leading to outcome misclassification), and chromosomal abnormalities are likely not related to the drug of interest.').

This study addresses all of these concerns. However, I had two main suggestions related to the analysis. One is to control the analyses for use of benzodiazepine and benzodiazepine-like drugs. if this is not a pre-planned analysis it could be reported as an exploratory secondary analysis. My rationale is that Oberlander et al. found an increased rate of malformations with SSRIs + benzodiazepine use but not SSRIs alone. Additionally, benzodiazepine exposure is fairly frequent in pregnancy and in a cohort with depression/anxiety diagnoses.

Response: We agree with Dr Mintzes. This was taken into account in all of our analyses by adjusting for co-medication use (including benzodiazepines). For clarity however, we are now presenting separate frequencies for benzodiazepine use in Table 1 and Table 2. Furthermore, this was pointed out in the Method section p. 11. We have also cited Oberlander et al. (ref. 30).

Indeed, co-medication during the first trimester was adjusted for (see p. 11, end of 1st paragraph). We have further presented co-medication use differently. We have separated benzodiazepines use from the other co-medications (Tables 1 and 2); ED, MD and hospitalizations are also presented more extensively in Tables 1 and 2. This did not change findings given that all these were already considered in the statistical analyses.

Secondly, given the large numbers and range of diagnoses included to construct this sub-cohort of pregnant women with a psychiatric diagnosis related to mood or anxiety, it would be helpful to include an additional adjustment for the severity level of the diagnosis, if only in broad categories, assuming a standardized scale or some guidance exists for such division. This again could be done as a secondary exploratory analysis. It would address the potential concern that within this large set of diagnoses, those with antidepressant exposure in pregnancy may be at a more severe illness level. Additionally, it would address the higher known smoking rates among patients with severe psychiatric illness. The lack of differences in psychiatric visits, except for TCA users and also the likelihood that some first trimester use is inadvertent ongoing use rather than a sign of severe illness, suggests that this may not be the case. However, it is worth carrying out an exploratory analysis to address this, given the large differences in severity of included diagnoses.

Response: Although we agree that residual confounding by severity of depression could remain in our study (acknowledged in the Discussion section – limitations), we do not have a specific measure of this in our cohort. Although many codes have been used to select our study cohort, the majority of our study population was depressed or anxious as expected given that they also had to be using antidepressant before pregnancy – this decreased indication heterogeneity, which explained why all were similar with regards to health care utilization (Tables 1 and 2). Nevertheless, our findings are

adjusted for markers of severity or of overall health such as co-medications, ED and hospitalizations, comorbidity, and SES. The fact that our findings are consistent with others published in general cohorts of pregnant women is reassuring and suggest that depression or markers of severity of depression are not important risk factors for major malformations.

The following sentence was added to the Discussion section on p. 17-18: 'Although residual confounding by severity of depression could remain in our study, our findings are consistent with others published in general cohorts of pregnant women, which is reassuring and suggesting that depression or markers of severity of depression are not important risk factors for major malformations.'

Related to this - is it possible to separate out ED visits, hospitalizations and psychiatric hospitalizations in the year pre-pregnancy, for the cohort as a whole? Given the frequency of hospitalizations or ED visits combined, this might provide more of a picture of how similar or different these groups were.

Response: As suggested by the Reviewer, ED, MD and hospitalizations are also presented more extensively in Tables 1 and 2. This did not change findings given that all these were already considered in the statistical analyses.

Some more minor points:
- please avoid the acronym MCM. As major congenital malformations are a key concept and this is not a general readership commonly used acronym, it should be spelled out in full for better readability;
Response : This was done as suggested throughout the revised manuscript.

- the wording 'were increasing the risk' or 'is increasing the risk' ' was statistically significantly increasing' should be changed, in part to improve the English ('were increasing' would be changed to 'increased') and in part for greater care to avoid claiming causation when an association has been observed; this is the case in the abstract and at various places in the text; this is the main reason I've noted a problem with the English; the English is fine otherwise.
Response: We thank the reviewer for pointing this out. We have modified the revised manuscript accordingly.

- given the references to the effects of amitriptyline and other TCAs on serotonin in the text and abstract, a reference is needed to support the statement in the text on effects on serotonin uptake;
Response: This has already been presented in the Introduction section on p. 6 (ref. 4): 'The mechanism of action of SNRIs is similar to SSRIs,4 and some TCAs, namely amitriptyline, also have serotonin inhibition effect.4'

- page 7, avoid the acronym "1DG"
Response : This was done as suggested throughout the revised manuscript.

- page 8, I was surprised at the decision to exclude pregnancies with multiple antidepressant exposures, rather than analyzing them as a separate category. It would be worth noting that total numbers were small, and therefore individual cells (e.g. specific types of combinations) were too small for analysis if this is the case.
Response : We thank Dr Mintzes for pointing this out. Indeed, there were very few combination/switches in our study cohort. We have added the following sentence for clarity as suggested in the Method section, p.8: '…..adverse pregnancy outcomes; this resulted in very few excluded pregnancies because the total number of combined AD uses or switches were small, and therefore individual cells (e.g. specific types of combinations) did not allow analyses….'.

- p11: 'the number of medication used other than antidepressants" should be "the number of medications..."

Response: Given that this included medication uses other than antidepressants, we reworded as follows in p. 11 (Method) and throughout the manuscript and tables: '….; the number of other medication uses including benzodiazepines;…..'

- Table 4: Given the large number of analyses presented in this table, it is likely to raise a concern that some associations would be expected by chance. One option in addressing this would be to add a column with a much tighter confidence interval than 95%. A second might be to distinguish between a small number of hypothesized associations based on the existing research on antidepressants and various forms of major malformations (such as paroxetine and cardiac malformations) and other exploratory analyses. This would then require adjustment for the secondary analyses. The table would need to be re-organized in this case, with hypothesized associations listed first, then additional organ systems below. This suggestion has implications for the related results reporting as well. Additionally, are not ventricular/atrial septal defects a subset of cardiac malformations? If the latter category includes overall cardiac malformations, this should be clear in the table design and a footnote. I did not understand the note that "results are provided when defects were identified for the antidepressant type." Does this mean that comparisons were only carried out if at least one user of a specific antidepressant had a child with the specific type of malformation?
Response: We have added the following footnote to Table 4 as suggested by the reviewer:
'Ventricular/atrial septal defects are a sub-set of cardiac malformations; Results are provided when defects were identified for the antidepressant type in the study population (if 0 outcome identified).'
With regards to multiple testing/comparisons: We agree that chance findings are always an issue in studies like ours. All authors (us and others) mentioned in our discussion section have handled the potential for chance finding the same way as we did. Indeed, we had already acknowledged the possibility of chance finding on p. 18, 1st paragraph. For clarity, we have added the underlined text below: 'In addition, given the number of comparisons made in this study, we cannot rule out the fact that some of our findings could have occurred by chance alone. It is however reassuring that many of the statistically significant associations that we found have been reported by other authors before, decreasing the possibility of chance finding.'
Literature on antidepressants in pregnancy is extensive. Some malformations can be identified for paroxetine but not for sertraline for example. It is thus difficult to present the findings in the order that Dr Mintzes suggested. We have however, clearly listed the malformations already identified in other studies in the revised Introduction (p. 6: 'Human studies concerning the adverse effects of antidepressant exposure during gestation on the developing fetus have showed increased risk of various congenital malformations such as cardiac, musculoskeletal, respiratory, cranyosynostosis, and cranyofacial,5-10 but variations between study results remain.11-15'; we are also comparing our findings with others on specific antidepressant types in the Discussion section.

- STROBE: "in pregnancy" is not the study design. The title does include the statement that this is a cohort study.
Response: We thank the reviewer for pointing this out. This has been corrected.

- Figure 1: The text in this Figure could be simplified by not repeating information in the reasons for exclusions and the inverse of the reason in the inclusion column. This might also allow for fewer boxes in the inclusion column, without information loss. Text could also be shortened per box or reason for exclusion.
Response: This was done as suggested.

- Figure 2: given that diagnoses in the year pre-pregnancy were measured during this entire period to construct the cohort and also that inadvertent use in early pregnancy and intended ongoing use during pregnancy among women with depression/anxiety diagnoses probably make up the large majority of first trimester exposures, rather than new depression/anxiety diagnoses during pregnancy, it would be useful to know whether there has been a growth in diagnoses of depression in the year

pre-pregnancy (or in the year pre-pregnancy and during pregnancy, in combination, if desired). This would be the more relevant information than only in-pregnancy diagnoses, as listed on Figure 2. Response: Figure 2 was made with data from the overall Quebec Pregnancy Cohort (and not only from the study population). With this, trends in maternal depression and gestational antidepressant use can be compared by calendar year in the QPC. The sentence addressing this on p. 11 of the Method section has been revised to: 'In addition, prevalence of maternal depression or anxiety, gestational use of antidepressants, and diagnoses of major malformations within the overall Quebec Pregnancy Cohort (and not only in the study population), stratified by calendar year of follow-up, were calculated; trend were tested using the Cochran-Armitage Trend test.' This gives us a sense of how many pregnant women were studied (how many were not eligible per design).

Please check the scale and data on this figure. I was surprised to see that the number of diagnoses in pregnancy exceeded antidepressant exposure in pregnancy. It is not clear if this is first trimester exposure only or exposure throughout pregnancy from the legend. Also the number reported as exposed in the first trimester in Figure 1 (36,440) represent 1.25% of the cohort of 289,688 after a number of exclusions. In Figure 2, the annual listed exposure rate grows from ~ 0.2% to 0.42%. This is inconsistent.
Response: Please see our response above.

The discussion mentions the concordance of the study results with other studies in administrative data that have examined rates of malformations among users of antidepressants, and specifically findings for paroxetine and cardiac malformations, citalopram and cranyosynostosis, and venlafaxine and respiratory defects. In addition to discussing the coherence of the current study with previous studies, it would be valuable to know how the effect size in the current study compared with those in studies that failed to restrict the comparison group to women with a mood or anxiety disorder diagnosis. If effect sizes are similar, this provides additional strength to the likelihood that there is little confounding by indication in relation to evidence of malformation rates with first trimester antidepressants. This finding would be especially important if it is possible to also control for broad categories of diagnoses, in order to add a severity component to the analysis, and for concurrent benzodiazepine use as recommended above.
Results: As suggested by the reviewer, we have added the effect sizes or effect size ranges of all of our comparisons with others. We have also added the following sentence on p. 16: 'This provides additional strength to the likelihood that there is little confounding by indication in relation to evidence of malformation prevalence with first trimester antidepressant use.'

As noted above, this is an important study that uses a large data set to evaluate malformation rates among pregnant women with antidepressant exposure in comparison with unexposed pregnant women with similar psychiatric diagnoses. It addresses an important question raised in this research domain concerning the influence of confounding by indication and unmeasured confounders on observed malformation rates, and adds considerably to the existing body of research evidence and our understanding of antidepressant effects.

VERSION 2 – REVIEW

REVIEWER	Angela Lupattelli University of Oslo, Norway
REVIEW RETURNED	04-Oct-2016

GENERAL COMMENTS	Thanks for the revised version of the manuscript. The authors have addressed most of the comments. In my view it is important and necessary to run sensitivity analyses

to address the robustness of the observed main findings. Whenever validation studies for exposure and/or outcome are available, parameters from these studies (sensitivity, specificity) can be used to explore the impact of exposure and outcome misclassification on the observed point estimates. The impact of specific unmeasured confounding can also explored via probabilistic bias analyses.

REVIEWER	BARBARA MINTZES THE UNIVERSITY OF SYDNEY
REVIEW RETURNED	24-Oct-2016

| GENERAL COMMENTS | I had wanted to re-check the latter, as I remembered a broad set of diagnoses being included under 'depression/anxiety'. This is important for the text in the article on diagnoses, and also stated prevalence rates for depression and anxiety. If a broader set of diagnoses were included that major depressive disorder and generalised anxiety disorder, this should be briefly described in the text, and rather than discussion depression prevalence, the authors should state 'prevalence of depression, anxiety disorders and related conditions'.

The rationale for using only depressed or anxious women without drug treatment as a comparison group is explained in the discussion. This should be stated when this comparison is first described, in the introduction or methods. My understanding is that there are two main rationales: 1) use of design to control for unmeasured confounders or for confounders not captured in this data set, such as consumption of alcohol, smoking rates, etc.(this is currently stated in the discussion); 2) because other results have been dismissed due to comparison with a general population of pregnant women, and this allows an assessment of whether the observed higher rate of malformations among women on antidepressants is robust.

The reason I suggest this early on is that the biological plausability for a very non-specific diagnosis of depression, anxiety or a range of related disorders to cause malformations is low. The rationale for the design is sound because of these two reasons but the reader should see this early on in the article.

A related point is that this design allowed the authors to compare their results to those of studies with a general population comparison group. They discuss this in a general way in 'discussion' but could be more systematic about this, as this is the most important added value of the current study. A table would be helpful comparing their estimated effect sizes to those of other authors, for major malformations in general and per type of malformation.

The issue of multiple comparisons could be better addressed for example by including two levels of significance, the second one adjusted for multiple comparisons (or presenting 99% as well as a 95% CI for outcomes). Some of their results would be robust to this, others not. Currently multiple comparisons is mentioned as a limitation but nothing is done to address this limitation although there are analytical solutions. Given the desire to take a precautionary approach in assessment of malformation rates with drug exposures in pregnancy, I agree with the decision to present all results that are significant at p = 0.05, but it would also be helpful to have a second |

analysis that highlights findings robust to the number of included comparisons. The other solution would have been a priori, to identify a small number of comparisons based on the existing research evidence that would be considered primary and therefore would not be adjusted for multiple comparisons but to adjust others that are more exploratory as there is no reason to expect that these might occur with a specific exposure.

I also wondered why the authors believe they have such a high observed rate of major malformation in this population as compared with historical norms. They report that 11% of unexposed pregnancies had major malformations. This is surprising if the usual rate is 3-5%.

Lastly, was any adjustment done for clustering of data by multiple pregnancies in the same woman.

A minor comment: there were some typos in the text, and some of the language could be improved. For example the construction 'drug x was increasing the rate of xx' should be replaced with 'drug x increased the rate'. And I understand that the unit of analysis is a pregnancy because one woman could be pregnant more than once, but I found it jarring to read about pregnancies rather than pregnant women because of the depersonalisation. The pregnant woman becomes a vessel rather than being seen as a person in her own right. I'd suggest putting the woman back in the picture.

There are also multiple references to results being reassuring because they are consistent with previous analyses. This should be rephrased to directly speak about consistency with previous analyses and not use the term reassuring. In several cases this is an increase in malformation rate - not at all reassuring for the families or infants.

Also in Table 1, under number of prescribers, I could not understand how a woman who was prescribed antidepressants could have 0 prescribers as she received at least one prescription.

The conclusion is confusing as it both states 'individualized treatment remains warranted' which suggests that despite the increased malformation risks an individual woman and provider may decide to use antidepressants in pregnancy but also fails to state what the basis might be for these individual decisions. The authors also refer to evidence of lack of effectiveness for mild to moderate depression. I thought they may have been hinting at the need for caution and to take the existence of alternative non-drug options and the variety of depression severities (and/or inclusion of many diagnoses for which there is no clinical trial evidence of antidepressant efficacy). However, they stop short of stating this clearly and the reader is left wondering what the key message is.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Angela Lupattelli
Institution and Country: University of Oslo, Norway
Competing Interests: None declared

Thanks for the revised version of the manuscript. The authors have addressed most of the comments.

In my view it is important and necessary to run sensitivity analyses to address the robustness of the observed main findings. Whenever validation studies for exposure and/or outcome are available, parameters from these studies (sensitivity, specificity) can be used to explore the impact of exposure and outcome misclassification on the observed point estimates. The impact of specific unmeasured confounding can also explored via probabilistic bias analyses.
Response : We have performed the sensitivity analyses as requested although we are mentionning in our Method section and in our Discussion section that both our data on exposures and malformations have been validated. We have revised as follows :

Method section on p. 13 : 'Although our data on medication exposures and congenital malformations have both been validated,17,21 we performed probabilistic sensitivity analyses (proposed by Lash and Fink23 and the SAS macro provided by Fox MP et al.24) to quantify the likely effects of misclassifications of exposure and outcome. For the exposure, we assumed a non-differential sensitivity analysis. The trapezoidal distributions for sensitivity and specificity with non-differential exposure misclassification were defined with a minimum of 75%, modes 85% and 95%, and a maximum 100%. For the analysis on the outcome misclassification, we used a differential sensitivity analysis and a separate trapezoidal distribution for exposed and unexposed pregnancies. Among exposed pregnancies, the sensitivity and the specificity were selected from trapezoid distributions with a minimum of 75%, modes 85% and 95% and a maximum of 100%. Among the unexposed pregnancies the trapezoidal distribution for sensitivity and specificity were defined with a minimum of 70%, modes 80% and 90% and a maximum of 95%. The correlation between the two sensitivity and specificity distributions was fixed at 0.8.

Result section, p. 16-17: 'Our sensitivity analyses on overall MCM have shown that our study results are robust (Table S8, Supplemental files). Indeed, taking into account potential misclassification of exposure and outcome, result in higher estimates of risk than what we have presented. Hence, our findings are conservative. Table S8 in the supplemental files show the non-differential probabilistic sensitivity analyses for antidepressant exposure misclassification; 95% simulation limits of 0.99 and 1.25, with a median (OR) of 1.11 for the conventional method, 1.09 to 4.11 with a median (OR) of 1.23 for the sensitivity analysis. Finally when the random error was added, the 95% simulation limits was 1.04 to 4.03 with a median (OR) of 1.25.
The differential probabilistic sensitivity analysis for major congenital malformation misclassification gave us a 95% simulation limit of 0.99 and 1.25, with a median (OR) of 1.11 for the conventional method; simulation limits of 0.43 to 55.15 with a median of 3.70 (OR) for the sensitivity analysis (Table S8, supplemental files). When the random error was added, the 95% simulation limits was 0.43 to 54.06 with a median (OR) of 3.72.'

Discussion section on p. 18: 'Although our data on antidepressant use and major malformations have been validated (ref, ref), our sensitivity analyses further show that our findings are valid and underestimate the true association.'

Given our results just presented above and the fact that our design adjusts for unmeasured confounders (as acknowledged by Reviewer 2 multiple times), we did not perform any probabilistic

analyses for the impact of unmeasured confounders. Using a cohort of depressed pregnant women indirectly adjusts for unmeasured confounders – we have already highlighted this in our Discussion section. This is true in our study and has been further showed in Huybrecht et al. (NEJM, 2016) – antidepressant users and non-users within a cohort of depressed pregnant women had similar distributions of smoking, alcohol use, etc.

………….
Reviewer: 2
Reviewer Name: BARBARA MINTZES
Institution and Country: THE UNIVERSITY OF SYDNEY, AUSTRALIA
Competing Interests: None

I had wanted to re-check the latter, as I remembered a broad set of diagnoses being included under 'depression/anxiety'. This is important for the text in the article on diagnoses, and also stated prevalence rates for depression and anxiety. If a broader set of diagnoses were included that major depressive disorder and generalised anxiety disorder, this should be briefly described in the text, and rather than discussion depression prevalence, the authors should state 'prevalence of depression, anxiety disorders and related conditions'.
Response : We thank the reviewer for this comment. We have presented all diagnostic codes used to define our study cohort in Supplemental file S1 (refered to in the Method section on p.7-8). Although it is true that many diagnoses were used, all study subjects either had depression or anxiety with or without other related disorders. This is now better explained on p.7-8 '…2) pregnancies with a diagnosis of depression and/or anxiety with or without concomitant related disorders…'. All the rational and discussion related to depression/anxiety remained the same given that they are still valid.

The rationale for using only depressed or anxious women without drug treatment as a comparison group is explained in the discussion. This should be stated when this comparison is first described, in the introduction or methods. My understanding is that there are two main rationales: 1) use of design to control for unmeasured confounders or for confounders not captured in this data set, such as consumption of alcohol, smoking rates, etc.(this is currently stated in the discussion); 2) because other results have been dismissed due to comparison with a general population of pregnant women, and this allows an assessment of whether the observed higher rate of malformations among women on antidepressants is robust.
The reason I suggest this early on is that the biological plausability for a very non-specific diagnosis of depression, anxiety or a range of related disorders to cause malformations is low. The rationale for the design is sound because of these two reasons but the reader should see this early on in the article.
Response : We thank the reviewer for this comment. We have revised as requested as follows :
p.6-7- Introduction : 'Because some findings in other studies have been dismissed due to comparisons with a general population of pregnant women, our design allowed us to assess whether the observed rate of malformations among women on antidepressants was robust.'
p. 9 – Method section : 'This allowed us to determine whether the observed rate of malformations among women on antidepressants was robust and independent of maternal depression status, and adjust for unmeasured confounders by design such as consumption of alcohol, smoking rates, and folic acid intake.'

The issue of multiple comparisons could be better addressed for example by including two levels of significance, the second one adjusted for multiple comparisons (or presenting 99% as well as a 95% CI for outcomes). Some of their results would be robust to this, others not. Currently multiple comparisons is mentioned as a limitation but nothing is done to address this limitation although there

are analytical solutions. Given the desire to take a precautionary approach in assessment of malformation rates with drug exposures in pregnancy, I agree with the decision to present all results that are significant at p = 0.05, but it would also be helpful to have a second analysis that highlights findings robust to the number of included comparisons. The other solution would have been a priori, to identify a small number of comparisons based on the existing research evidence that would be considered primary and therefore would not be adjusted for multiple comparisons but to adjust others that are more exploratory as there is no reason to expect that these might occur with a specific exposure.

Response : As requested, we have added a 99% CI in Supplemental file S7. All findings on citalopram remained statistically significant, which means that our estimates are robusts. However, findings on paroxetine were non-significant when using a 99% CI. This in itself is not an issue given that our findings are replicating findings from other studies and meta-analyses. This has been presented in the :

Method section, p. 12 : 'Further 99%CIs were calculated to determine the robustness of the findings.'
Result section, p. 16: 'Table S7 presents organ specific estimates using 99%CI to assess robustness of estimates. Results on citalopram remained statistically significant. However, all other findings were non-statistically significant using 99%CIs.'
Discussion section, p.20 : 'Furthermore, findings on citalopram are robust as shown when calculating wider confidence intervals (99%CI). This is less of an issue for paroxetine given that findings from our study have replicated estimates already published in other studies.'

I also wondered why the authors believe they have such a high observed rate of major malformation in this population as compared with historical norms. They report that 11% of unexposed pregnancies had major malformations. This is surprising if the usual rate is 3-5%.
Response : We agree with the reviewer but this has been described before when using data from the QPC cohort. For clarity and precision we have added the following in the Discussion section on p.21 : 'Our MCM population prevalence may seem somewhat higher than the routinely reported 3-5%, but in fact our rate is consistent with what is expected in the province of Quebec, due to high concentration of genetic risk factors stemming from the 'founding' French ancestors.43 Nevertheless, given that the baseline rate of MCM is similarly higher in those taking antidepressants vs non-users in our study, it has no impact on the internal validity of our study because it cancels out when comparing the exposed and unexposed groups.'

Lastly, was any adjustment done for clustering of data by multiple pregnancies in the same woman.
Response : Yes. The use of GEE models adjust for multiple pregnancies per women during the follow-up period. For clarity, we have added the following sentences in the Method section, p. 12 : 'Crude and adjusted odds ratios (OR) with 95% confidence intervals (95%CI) were calculated for each outcome separately using generalized estimating equation (GEE) models, which take into account multiple pregnancies per women during the follow-up period. Further, 99%CIs were calculated using GEE models to determine the robustness of the findings.'

A minor comment: there were some typos in the text, and some of the language could be improved. For example the construction 'drug x was increasing the rate of xx' should be replaced with 'drug x increased the rate'. And I understand that the unit of analysis is a pregnancy because one woman could be pregnant more than once, but I found it jarring to read about pregnancies rather than pregnant women because of the depersonalisation. The pregnant woman becomes a vessel rather than being seen as a person in her own right. I'd suggest putting the woman back in the picture.
Response : We thank the reviewer and have revised the manuscript accordingly. With regards to the comment on ' was increasing' vs. 'increased'. These changes were done following Reviewer 1's comments in the first revision of the manuscript.

There are also multiple references to results being reassuring because they are consistent with

previous analyses. This should be rephrased to directly speak about consistency with previous analyses and not use the term reassuring. In several cases this is an increase in malformation rate - not at all reassuring for the families or infants.
Response : We thank the reviewer again and have revised the manuscript accordingly

Also in Table 1, under number of prescribers, I could not understand how a woman who was prescribed antidepressants could have 0 prescribers as she received at least one prescription.
Response : We thank the reviewer for pointing this out. There was a label typo. It should read : 0-1, 2-3, >3. This has been corrected in Table 1 and 2.

The conclusion is confusing as it both states 'individualized treatment remains warranted' which suggests that despite the increased malformation risks an individual woman and provider may decide to use antidepressants in pregnancy but also fails to state what the basis might be for these individual decisions. The authors also refer to evidence of lack of effectiveness for mild to moderate depression. I thought they may have been hinting at the need for caution and to take the existence of alternative non-drug options and the variety of depression severities (and/or inclusion of many diagnoses for which there is no clinical trial evidence of antidepressant efficacy). However, they stop short of stating this clearly and the reader is left wondering what the key message is.
Response : We thank the reviewer for this comment. We have revised our Conclusion on p. 22 as follows : 'In this population-wide cohort study, we found that infants were at an increased risk of cardiac, musculoskeletal, craniofacial, digestive and respiratory defects as well as craniosynostosis from in-utero exposure to serotonin inhibitor drugs (SSRI, SNRI and some TCAs). Given that an increasing number of women are diagnosed with depression during pregnancy, these results have direct implications on their clinical management. This is even more important given that the effectiveness of antidepressants during pregnancy for the treatment of the majority of cases of depression (mild to moderate depression) have been shown to be marginal.42 Hence, the need for caution with antidepressant use during pregnancy is warranted and alternative non-drug options should be considered.'