PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Low alcohol consumption and pregnancy and childhood outcomes: time to change guidelines indicating apparently 'safe' levels of alcohol during pregnancy? A systematic review and meta-analyses.
AUTHORS	Mamluk, Loubaba; Edwards, Hannah; Savovic, Jelena; Leach, Verity; Jones, Tim; Moore, Theresa; Ijaz, Sharea; Lewis, Sarah; Donovan, Jenny; Lawlor, Debbie; Davey Smith, George; Fraser, Abigail; Zuccolo, Luisa

VERSION 1 - REVIEW

REVIEWER	Darren Greenwood University of Leeds
REVIEW RETURNED	08-Dec-2016

GENERAL COMMENTS	General: A really interesting research question, with potential high impact. And it's topical, given that even quite recently DoH, NICE and RCOG have made what the public might consider as inconsistent suggestions on the number of alcohol units that are safe for pregnant women to consume. A detailed review of the evidence could therefore make some useful contribution to the debate. Major points: 1. It is disappointing that the specific cut-off value of 32g/day has not been evaluated. Surely that is the main item of interest? In addition, potentially greater power could be drawn from investigating any dose-response. This could also contribute to discussion of possible causality. 2. The paper would benefit from much more assessment of risk of bias *within* studies. It is usual practice to have a separate table assessing this, often using some standardized method such as the Newcastle-Ottawa score (though reporting each dimension separately rather than summing to an overall score). This forms an important component of the PRISMA AND MOOSE guidelines. 3. Given the heterogeneity between these observational studies, the paper would benefit from much more exploration of that heterogeneity. This could provide some additional clinical insight that could inform future studies or explain different results from different studies. This forms an important component of the MOOSE guidelines. Minor points: 4. It is nice to have the search no more than 3 months out of date on submission, to maintain relevancy on publication. This one is up to around 5 months ago, so if any substantive corrections are recommended, it might be worth re-running the searches to rule out any new studies published in the meantime.

	5. Text sometimes says things like “estimates for preterm birth included the null value…”. I think this refers to the confidence including the null value. However, the point is not so much this, but more generally that the confidence interval includes quite a wide range of possible associations, including no association, but also the possibility of an almost 30% increase in odds. So maybe the headline here is that we still don’t know, because there’s not much evidence out there, which is pretty much where we started. As such, the review doesn’t shed as much new light on the topic as we might have hoped. 6. It is unusual to include different study designs, e.g. cohorts, quasi-experimental studies, negative control studies, or Mendelian Randomization in the same meta-analysis, because this introduces a known element of unnecessary heterogeneity. Sometimes this can’t be avoided because there are too few of each type, but here most of the information comes from cohorts, so for me it would be better to only include cohorts in meta-analysis, but incorporate the other study designs in the wider evidence synthesis. The MR study is well worth acknowledging in its own right, for example. As far as I can tell, the different study designs have been separated, but this is worth clarifying in the methods (sorry if I’ve missed this). 7. Given the uncertainty in the estimated heterogeneity, it is over-precise to quote the I-squared values to anything more than the nearest integer. 8. Page 9, line 37 “Where appropriate, we additionally pooled results for each outcome”. Please state how. 9. Page 9, line 41 “When continuous outcomes were measured using different scales…” please indicate which outcomes this refers to. 10. Page 9, line 47 “results were pooled unless they were very different from each other (I2 ≥ 50%)”. Yet pooled estimates are quoted with I-squared more than this. Please clarify. 11. Page 10, line 7 “likelihood of small study bias deriving from publication bias was assessed through visual inspection of funnel plots…”. This would be more clearly referred to as small study bias “such as” publication bias. The source cannot be known. 12. Page 10, line 9 “inspection of funnel plots for pooled analyses including ≥4 studies”. General guidelines are that we cannot assess funnel plot asymmetry with fewer than 10 studies. Please clarify. 13. The Nykjaer et al 2014 is not the “UK Women’s Cohort” (if middle-aged women) or the “UK cohort for birthweight” as inconsistently described, but the “Caffeine and Reproductive Health (CARE) Study” as they state and as described elsewhere in the current manuscript. 14. Labelling of the horizontal axes on the forest plots needs improving. Use of round numbers is preferred. 15. The stratified forest plots in the supplementary material would benefit from better labelling of the subgroups.

REVIEWER	Dr Lesley Smith Oxford Brookes University, UK
REVIEW RETURNED	23-Jan-2017

GENERAL COMMENTS	This systematic review and meta-analysis has been conducted to a high standard but I have some major comments which are listed below. Abstract: The results could be made clearer by indicating whether

	the ORs are adjusted or unadjusted and what factors were adjusted for. Also giving some indication of number of studies/participants in each of the estimates as is currently mis-leading – 22 studies in the review but not in each meta-analysis. The timepoint of exposure should also be made clear. Outcomes page 7 – lines 3-7: some of the outcomes need defining as they can vary from study to study and for readers unfamiliar with obstetric outcomes e.g. SGA, pre-term delivery or outcomes presented in the meta-analayses. Page 8, lines 13-18: Known confounders – this seems like a minimum list so some justification for excluding other confounders would be useful e.g. parity. Also elaboration of what aspects of socioeconomic positioning you considered would be useful. This is important to clarify as you cite this as a strength of the paper. Page 8, line 32: please clarify…'reference group, categories of exposure' as unclear what they refer to and how they are different. Data analysis Page 9, line 23: please justify why you chose to include estimates relative to the earliest exposure if exposure during different trimesters were presented in a paper. Figure 1 and values in text do not add up – 784-763 studies = 22 studies? Page 10, standard analytic approaches: The results could be presented in a more logical order. Figure 3A presented before fig 2 results. Line 37 – error in number of studies – should be 6? In addition to reporting the number of studies in each meta-analysis, the event rate is important too for the binary outcomes. Line 52: typo for birthweight results for summary effect and CI values. Figure 3c results not presented? Page 11: I would query whether 4 studies as a cut-off in a Funnel plot is enough to say anything about small study bias. Page 13, Table 1: A lot of information in the Table seems redundant as is also in the Forest plots, and rather than knowing whether an analysis was adjusted or not, knowing the factors adjusted for would be more useful. The number in analyses does not report the number with the outcome in each group (event rate) which would be useful to help interpretation of findings. Rather than labeling each study with the first author of a paper, the cohort name would be useful as many are in common usage and widely recognized. Table 2: I would find this table more useful if the studies were sorted in a logical presentation order. Perhaps by outcome and show the studies reporting on the outcomes in the meta-analyses first, then for the outcomes for which there were fewer studies. Or by pregnancy, then infant developmental outcomes. Perhaps break it up a bit and use some sub-headings. Help the reader navigate this large table of data. Discussion: Answer you research question rather than re-stating it in the first line of the discussion.

	I would switch the order of (i) and (ii). Strengths and limitations, page 18: If the results for the alternative study designs are presented in another review then I don't think you can claim that the effort to include them is a strength in this review. You need to expand on why you think residual confounding due to SE position is an issue given that you claim it is a strength that you included studies that adjusted for SE position? A major limitation is the lack of information around timing of exposure. The distinction between a reporting bias of the review or individual studies is not clear. Page 20, line 10. The proportion of women drinking during pregnancy (up to 80%) needs context – is this any drinking at any point i.e. light and occasional. Elaborate on this point. Page 20. Please give some indication and discussion around the clinical significance of the effect sizes reported in the paper. Conclusion, line 40-42: given the review has assessed light drinking (in line with pregnancy advice), it isn't clear to me how these findings will help women anxious about the effects of drinking before pregnancy recognition. Drinking before pregnancy recognition is typically higher than the level assessed in this review. Please remove this sentence.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Darren Greenwood
Institution and Country: University of Leeds
Please state any competing interests: None declared

General:
A really interesting research question, with potential high impact. And it's topical, given that even quite recently DoH, NICE and RCOG have made what the public might consider as inconsistent suggestions on the number of alcohol units that are safe for pregnant women to consume. A detailed review of the evidence could therefore make some useful contribution to the debate.

Major points:

1. It is disappointing that the specific cut-off value of 32g/day has not been evaluated. Surely that is the main item of interest? In addition, potentially greater power could be drawn from investigating any dose-response. This could also contribute to discussion of possible causality.

Response: this is an important point. We recognise the importance of this however with the small amount of studies having evaluated the effect of 32g/week, it was not possible to investigate a dose-response within the 0-32 g/week range. The 32 g/week dose was chosen as the relevant comparison category because this was the 'safe' cut-off stated in most UK guidelines (nice, doh, rcog) at the time of designing this review, and indeed the aim was to compare this 'safe' dose to an alternative recommendation of 0 g/week.
A further piece of work could indeed address this issue of whether there is a threshold effect at

32g/week. However, given the scarse evidence base, we doubt that this would give conclusive answers.

2. The paper would benefit from much more assessment of risk of bias *within* studies. It is usual practice to have a separate table assessing this, often using some standardized method such as the Newcastle-Ottawa score (though reporting each dimension separately rather than summing to an overall score). This forms an important component of the PRISMA AND MOOSE guidelines.

Response: the authors acknowledge the importance of this and have now carried out a risk of bias assessment using the newcastle-ottawa tool as suggested by the reviewer. (pages 10 & 11 and supplementary Table 3).

3. Given the heterogeneity between these observational studies, the paper would benefit from much more exploration of that heterogeneity. This could provide some additional clinical insight that could inform future studies or explain different results from different studies. This forms an important component of the MOOSE guidelines.

Response: thank you for this comment. Only one outcome has an I2 60% (preterm birth). All studies used the same definition of the outcome and control for similar confounders. All studies also have the same study design. We decided against running a formal meta-regression due to the small number of studies, substantially limiting our statistical power to identify study-level effects (Higgins jp, green s, editors. Cochrane handbook for systematic reviews of interventions. John Wiley & sons; 2011 aug 24).
This has now been added:
“additionally, most studies assessing preterm birth had corrected for main confounders know to be associated with preterm birth, with the exception of one study [15] that did not correct for any”. (page 12)

Minor points:
4. It is nice to have the search no more than 3 months out of date on submission, to maintain relevancy on publication. This one is up to around 5 months ago, so if any substantive corrections are recommended, it might be worth re-running the searches to rule out any new studies published in the meantime.

Response: no substantive corrections were needed and we returned our revised manuscript within one month of receiving the referees' comments, thus minimising delays. We have therefore not updated the search.

5. Text sometimes says things like “estimates for preterm birth included the null value…”. I think this refers to the confidence including the null value. However, the point is not so much this, but more generally that the confidence interval includes quite a wide range of possible associations, including no association, but also the possibility of an almost 30% increase in odds. So maybe the headline here is that we still don’t know, because there’s not much evidence out there, which is pretty much where we started. As such, the review doesn’t shed as much new light on the topic as we might have hoped.

Response: the only instance using such wording is in the abstract where numerical results are also provided alongside the text and we have not changed this, in order to comply with the abstract word limit. We would also argue that the value of this and indeed any review is to highlight not only what is already known but also where important knowledge gaps still exist.

6. It is unusual to include different study designs, e.g. cohorts, quasi-experimental studies, negative

control studies, or Mendelian Randomization in the same meta-analysis, because this introduces a known element of unnecessary heterogeneity. Sometimes this can't be avoided because there are too few of each type, but here most of the information comes from cohorts, so for me it would be better to only include cohorts in meta-analysis, but incorporate the other study designs in the wider evidence synthesis. The MR study is well worth acknowledging in its own right, for example. As far as I can tell, the different study designs have been separated, but this is worth clarifying in the methods (sorry if I've missed this).

Response: the authors agree with the reviewer and as per protocol, we have not pooled these together. This review analysed the different study designs separately. We have now added the following sentence to add clarification.
"results from different study designs have been reviewed separately" (page 9)

7. Given the uncertainty in the estimated heterogeneity, it is over-precise to quote the I-squared values to anything more than the nearest integer.

Response: this has now been changed to (I2=60%) (page 12) and abstract (page 3)

8. Page 9, line 37 "Where appropriate, we additionally pooled results for each outcome". Please state how.

Response: if the results could be pooled and combined with at least one other result from a different study, then they were included in the meta-analysis of a specific outcome. This was done using random effects meta-analysis as stated in the methods section (page 9)

9. Page 9, line 41 "When continuous outcomes were measured using different scales…" please indicate which outcomes this refers to.
We thank the reviewer for drawing our attention to this.

Response: this sentence refers to our original intention should we find studies measuring continuous outcomes on different likert-scale (e.g. cognitive outcomes). However, this actually did not materialise and we have therefore removed this from the manuscript (page 9) but of course this remains in the protocol.

10. Page 9, line 47 "results were pooled unless they were very different from each other (I2 ≥ 50%)". Yet pooled estimates are quoted with I-squared more than this. Please clarify.

Response: this reviewer is correct, however this sentence is preceded by the following sentence "where only two studies were available to meta-analyse, results were pooled unless they were very different from each other (I2 ≥ 50%)" indicating that this was the case when there were only 2 studies.

11. Page 10, line 7 "likelihood of small study bias deriving from publication bias was assessed through visual inspection of funnel plots…". This would be more clearly referred to as small study bias "such as" publication bias. The source cannot be known.

Response: this has now been changed to "The likelihood of small study bias, such as publication bias, could not be assessed through visual inspection of funnel plots for pooled analyses as no outcome was assessed by 10+ studies [18]." (page 10)

12. Page 10, line 9 "inspection of funnel plots for pooled analyses including ≥4 studies". General guidelines are that we cannot assess funnel plot asymmetry with fewer than 10 studies. Please clarify.

Response: we accept this comment and have therefore deleted the funnel plots and all reference to them.

13. The Nykjaer et al 2014 is not the “UK Women’s Cohort” (if middle-aged women) or the “UK cohort for birthweight” as inconsistently described, but the “Caffeine and Reproductive Health (CARE) Study” as they state and as described elsewhere in the current manuscript.

Response: thank you for this correction. This has now been corrected in figure 2 and figure 3c.

14. Labelling of the horizontal axes on the forest plots needs improving. Use of round numbers is preferred.

Response: thank you. These have now been rounded in all plots.

15. The stratified forest plots in the supplementary material would benefit from better labelling of the subgroups.

Response: thank you for this. This has now been done (page 32).

Reviewer: 2
Reviewer Name: Dr Lesley Smith
Institution and Country: Oxford Brookes University, UK
Please state any competing interests: None declared

Please leave your comments for the authors below

This systematic review and meta-analysis has been conducted to a high standard but I have some major comments which are listed below.

Abstract: The results could be made clearer by indicating whether the ORs are adjusted or unadjusted and what factors were adjusted for. Also giving some indication of number of studies/participants in each of the estimates as is currently mis-leading – 22 studies in the review but not in each meta-analysis. The time point of exposure should also be made clear.

Response: this section now reads:

“summary odd ratios (OR) 1·08, 95% confidence intervals (CI) (1·02 to 1·14), I2 0%, (7 studies, all estimates were adjusted) or 1·10, 95% CI (0·95 to1·28), I2 60%, (9 studies, includes one unadjusted estimates) respectively. The earliest time points of exposure were used in the analysis” (page 3)

Outcomes page 7 – lines 3-7: some of the outcomes need defining as they can vary from study to study and for readers unfamiliar with obstetric outcomes e.g. SGA, pre-term delivery or outcomes presented in the meta-analyses.

Response: thank you. This section now reads:
“Outcomes included: 1) pregnancy outcomes: still birth (pregnancy loss after week 24; miscarriage; gestational length and preterm delivery (<37 weeks gestation); hypertensive disorders of pregnancy; gestational diabetes; small for gestational age (SGA, < 10th percentile in weight or <−2 standard deviation scores) and birth size (weight (including low birth weight defined as <2500g), length, and head circumference); low amniotic fluid (oligohydramnios);
We also added to the next paragraph “We adopted study specific definitions for all outcomes” (page 7).

Page 8, lines 13-18: Known confounders – this seems like a minimum list so some justification for excluding other confounders would be useful e.g. parity. Also elaboration of what aspects of socioeconomic positioning you considered would be useful. This is important to clarify as you cite this as a strength of the paper.

Response: thank you for this comment. These variables were chosen a priori based on existing knowledge about their relationship with both the exposure of interest and the outcomes. We took a pragmatic approach and considered any measure of SEP to be a potential confounder and have now added this to the manuscript for clarity.
"We assessed potential for bias in included studies by assessing how well the study adjusted for several main confounders known to impact on the exposure-outcome associations (socioeconomic positioning as measured by the individual study, smoking during pregnancy, maternal age, and ethnicity)." (page 8)

Page 8, line 32: please clarify…'reference group, categories of exposure' as unclear what they refer to and how they are different.

Response: thank you for pointing this out. Throughout the paper, we always mean reference group= mothers consuming no alcohol during pregnancy.
This now reads "refence group (abstinence), exposure (e.g 1-2 units or 2-4 units)".
(page 8)
Data analysis Page 9, line 23: please justify why you chose to include estimates relative to the earliest exposure if exposure during different trimesters were presented in a paper.

Response: this has now been added:
This is because for some outcomes, the first trimester tends to be the most critical timing/window of exposure [12] [13] and because most studies that only reported on one time point reported on exposure in early gestation. (page 9)

Figure 1 and values in text do not add up – 784-763 studies = 22 studies?

Response: thank you for pointing this out. This has now been changed (page 29)

Page 10, standard analytic approaches: The results could be presented in a more logical order. Figure 3A presented before fig 2 results.

Response: this now no longer follows order by number of studies and reads:
Figure 2 presents results for birthweight (7 studies). Figure 3A presents results for preterm delivery (9 studies) Figure 3B, presents results for SGA (7 studies), and results for low birthweight (6 studies) are given in Figure 3C. (page 12)

Line 37 – error in number of studies – should be 6?

Response: thank you for pointing this out. This has now been changed to 6 studies for low birthweight.

In addition to reporting the number of studies in each meta-analysis, the event rate is important too for the binary outcomes.

Response: this has now been added in the plots when that information was available. When studies did not report the actual number of events we left this blank.

Line 52: typo for birthweight results for summary effect and CI values.
Response: thank you for pointing this out. This has now been corrected “summary effect of -13·49g (95% CI -30·28g; +3·31g)” (page 12)
Figure 3c results not presented?

Response: thank you for this. It has now been added:
“summary effects for birthweight <2500g were, or 1·00, 95% CI 0·82; 1·22 (figure 3c)” (page 12)

Page 11: I would query whether 4 studies as a cut-off in a Funnel plot is enough to say anything about small study bias.

Response: we accept this comment and have therefore deleted the funnel plots and all reference to them, as noted above.

Page 13, Table 1: A lot of information in the Table seems redundant as is also in the Forest plots, and rather than knowing whether an analysis was adjusted or not, knowing the factors adjusted for would be more useful.

Response: all adjustments for each study are in the supplementary Table 2 (page 39).

The number in analyses does not report the number with the outcome in each group (event rate) which would be useful to help interpretation of findings.

Response: these have now been added in Table 1 and in the plots where relevant and where information was available.

Rather than labelling each study with the first author of a paper, the cohort name would be useful as many are in common usage and widely recognized.

Response: thank you for your comment. The cohort names are labelled in the plots.

Table 2: I would find this table more useful if the studies were sorted in a logical presentation order. Perhaps by outcome and show the studies reporting on the outcomes in the meta-analyses first, then for the outcomes for which there were fewer studies. Or by pregnancy, then infant developmental outcomes. Perhaps break it up a bit and use some sub-headings. Help the reader navigate this large table of data.

Response: we have now grouped the studies by outcome as described in the methods section as advised, thank you. (page 17- 19).

We would like to clarify that table 2 only includes studies that have not been pooled in the meta-analysis.

These studies are listed by publication date within each outcome.

Discussion: Answer you research question rather than re-stating it in the first line of the discussion.
I would switch the order of (i) and (ii).

Response: thank you. We prefer to keep the order of those two sentences, instead, we changed the first sentence to this:
"in this comprehensive systematic review of the literature on the effects of low levels of alcohol

drinking in pregnancy, the two main findings are:"
(page 18)

Strengths and limitations, page 18: If the results for the alternative study designs are presented in another review then I don't think you can claim that the effort to include them is a strength in this review.

Response: we have included results from 2 negative control studies in the present review, those reporting on outcomes comparing mothers consuming up to 32g alcohol/week versus none. Results from other alternative study designs that did not fit with our strict dose-specific inclusion criteria are not reported here. We maintain that the inclusion of studies using an analytical approach that minimises bias and/or confounding is a strength.

You need to expand on why you think residual confounding due to SE position is an issue given that you claim it is a strength that you included studies that adjusted for SE position?

Response: thank you for your comment. This has now been added:

SE position is a complex, multi-faceted entity. Several studies have attempted to adjust for se position by collecting information on, for example, maternal education, family-level SE position around the time of the pregnancy, home address-based deprivation index etc. Few studies included more than one of these measured (17 24 25 26). Whereas we consider attempting to adjust for at least one of these characteristics to be a minimum requirement to account for some of the confounding introduced by SE position, there remains scope for residual confounding. [48] Given the strong relationship between SE position and both the exposure (alcohol use in pregnancy) and outcomes in this review, any degree of residual confounding is of course an issue when interpreting the effect estimates from the observational studies included in this review. (page 21 and 22)

A major limitation is the lack of information around timing of exposure. The distinction between a reporting bias of the review or individual studies is not clear.

Response: this has now been added:

"this also was the case for identifying effects based on time of exposure, which is also a limitation." (page 22)

We have now added a column in Table 1 regarding time of exposure. (page 16)

Page 20, line 10. The proportion of women drinking during pregnancy (up to 80%) needs context – is this any drinking at any point i.e. light and occasional. Elaborate on this point.

Response: this has now been changed to:

"…with up to 80% of women consuming some alcohol during pregnancy." (page 23)

Page 20. Please give some indication and discussion around the clinical significance of the effect sizes reported in the paper.

Response: We have added the following to provide clinical context to our findings:

Here we found that maternal alcohol consumption of up to 32g/week was associated with an 10%

increased risk of preterm birth (95%CI: 0·95 to1·28). In comparison, light to moderate smoking (<20 cigarettes per day) is associated with a 22% increased risk of preterm birth (95% CI: 1·13 to 1·32). [61] (page 23)

Conclusion, line 40-42: given the review has assessed light drinking (in line with pregnancy advice), it isn't clear to me how these findings will help women anxious about the effects of drinking before pregnancy recognition. Drinking before pregnancy recognition is typically higher than the level assessed in this review. Please remove this sentence.

Response: this sentence has now been removed.

VERSION 2 – REVIEW

REVIEWER	Dr Darren Greenwood University of Leeds, UK
REVIEW RETURNED	25-Apr-2017

GENERAL COMMENTS	The authors have provided a clear and helpful response to the points raised. I am happy that all my main concerns have been adequately addressed and only the following minor point remains, which is not a deal-breaker: 5. I accept that this is a minor presentational point, but reference to the "null" in the abstract is rather mathematical language for a general medical journal. I still think it would be clearer to the general readership to talk about the possibility of no association remaining, rather than referring to some mathematical null hypothesis that has not even been stated. I don't think it would take many more words.

VERSION 2 – AUTHOR RESPONSE

Thank you very much for recommending our manuscript for publication.
Please find the response to the reviewer's comments below.

Reviewer: I accept that this is a minor presentational point, but reference to the "null" in the abstract is rather mathematical language for a general medical journal. I still think it would be clearer to the general readership to talk about the possibility of no association remaining, rather than referring to some mathematical null hypothesis that has not even been stated. I don't think it would take many more words.

Response: thank you for your comment. This has now been changed to the following.

Odds of small-for-gestational-age (SGA) and preterm birth were higher for babies whose mothers consumed up to 32g/week versus none, but estimates for preterm birth were also compatible with no association. (page 3)