RESPONSE TO REVIEWERS' COMMENTS

Manuscript ID BMJ.2017.039236.R1 titled " Impacts of London's road traffic air and noise pollution on birth weight: a retrospective population-based cohort study."

We would like to say how much we appreciate the time taken by the Reviewers, once again, and the feedback they have provided. We have addressed the Reviewers' comments point by point below. We provide clean and tracked changes versions of the revised manuscript. We have additionally separated out the Supplemental Material into a separate document now – but this is also provided in a tracked changes and clean version. Our responses below are prefaced by "Authors' Response" and shown in blue to distinguish from the Reviewers' comments. Line numbers we mention in responses refer to manuscript version with tracked changes.

Reviewer: 1 - Ulrike Gehring

Comments:
The revision of this manuscript includes many substantial changes that elevate its quality. I have some remaining comments and suggestions.

Comments
Abstract, conclusions: Conclusions are too long. They should be 1 or 2 sentences, not more. Sentence about attributable risk should be moved to the results section. In the light of the associations with non-exhaust air pollution, the reference to diesel should be removed from the abstract.

Authors' Response: We have shortened the Abstract Conclusions as suggested by the reviewer, i.e. moved attributable risk to Results (line 48), and deleted the mention of diesel.

Study population. I agree with the removal of outcomes in all birth and the restriction of the paper to term births as this makes the paper easier to follow. However, the paper in its current form is not consistent with regard to the inclusion/exclusion preterm birth, for example the study population and Table 1 should still include preterm births. Moreover, line 220 mentions stratified analyses by term/preterm. That does not make sense if all analyses were performed in term births. This needs to be fixed.

Authors' Response: Thank you for flagging this up. We have rectified this throughout the manuscript. Study population in abstract (line 32), Table 1 (and Supplementary Tables 1, 2, and 3) revised to exclude preterm births. We have modified text re 'exclusions' at line 183 to remove preterm births. We have deleted at line 221 the reference to stratified analyses by term/preterm.

Methods, line 127-128. The use of the air pollution model for other epidemiological analyses with other health endpoints does not say anything about the model's validity. Therefore, this sentence including the 5 references should be removed from the paper.

Authors' Response: We have deleted this sentence at line 126 as suggested by the reviewer.

Line 156f. Randomly distributed across what? Could you please clarify that.

Authors' Response: randomly distributed across the study area. We have amended at line 155 to clarify this.

Use of noise as a categorical variable. The plots of the GAMs included in the reply and should be included in the (supplement of) the paper. The plots indicate that outcome-noise relationships do not severely deviate from linearity. I think the reasons for continuing to use noise as a categorical variable do not outweigh the limitations of such an analysis, i.e. that the use of air pollution as continuous variables and noise as categorical variables makes it difficult to compare effects and (as mentioned by reviewer 3) and to determine the impact of mutual adjustment. Moreover, the non-normal distribution of the exposure variable is not a requirement in linear (logistic) regression – of course, when using non-transformed noise in linear/logistic regression needs careful checking of associations being driven by extremely high exposures. In the light of the above, I suggest to replace associations with categorical noise variables by associations with continuous noise variables expressed per IQR increase in exposure, or at least to add associations with continuous noise variables to the paper. Moreover, GAMs assessing the functional relationship between outcomes and exposure should be introduced before linear (logistic) regression results.

Authors' Response:
We have added the GAMs included in our previous response document to the Supplemental Material (now Supplementary Figure 1) as suggested by the Reviewer, and mentioned these in the text of the manuscript at line 199.

We stand by our decision to analyse noise as a categorical variable, as per our previous response this was not solely determined by the noise distribution but also to reduce potential exposure misclassification (see lines 395-398 of the Discussion).

However, in line with the Reviewer's suggestion, we have added as Supplementary Table 8 the results of noise analysed on a continuous scale (per IQR), for the full range of linear models i.e. unadjusted and adjusted single-exposure models, and joint air pollution-noise models.

In the joint air pollutant-noise models which included both air pollutants and noise as a continuous variable analysed on the IQR scale:

- There was no association between either night-time or day-time noise and term LBW or term SGA.
- There was no association between either night-time or day-time noise and mean term birth weight in models adjusted for primary traffic-related air pollutants (NO_2, NO_X, $PM_{2.5\ traffic\text{-}exhaust}$, $PM_{2.5\ traffic\text{-}non\text{-}exhaust}$), although the coefficient in the models adjusted for NO_2 and NO_X was negative (but the CI included 1).
- The associations between air pollutants and term LBW, term SGA and mean term birth weight were virtually identical to those from models which included noise as a categorical variable. The coefficients for mean term birth weight differed by less than 1g between the models.

In short, these results are consistent with the joint air pollutant-noise models which included air pollutants as continuous variables and noise as categorical variables, and do not change the conclusions of the paper. We have added some text to the Methods section at line 202 to reflect these analyses and have integrated these analyses into the Results section between lines 284-304 .

We have integrated better the various analyses (noise as categorical, noise as continuous, and GAM models) throughout the Results section (lines 249-313).

Discussion, line 401. Add "vehicles" after 10,000.

Authors' Response: thanks for pointing this out. We have amended as suggested at line 378.

Lines 430ff. You state that you did not have data on ambient temperature. I assume that routine data is available for the area and period of interest. Do you mean that the temperature data have never been linked to the birth weight data and cannot be linked at this stage? Please clarify.

Authors' Response: That is correct, temperature data have never been linked to the birth weight data and cannot be linked at this stage. Amended at line 415 to read "we did not have data on temperature linked to the births data".

Line 447, multiple testing. This is very short. I suggest to add why you decided not to adjust for multiple testing despite the limitations of doing so.

Authors' Response: We have acknowledged the multiple testing issue and stated whether or not we have made adjustment for this (and we have now additionally stated this in the Methods at line 231, for the avoidance of any doubt). This is in line with "Statistical Analyses and Methods in the Published Literature (SAMPL) Guidelines" (http://www.equator-network.org/wp-content/uploads/2013/03/SAMPL-Guidelines-3-13-13.pdf), which are referred to in the BMJ Guidelines to Authors as the guidelines for reporting statistical aspects of the study. We did not adjust for multiple testing as the commonly used method (Bonferroni) is too conservative. We acknowledge that there are more complex methods available, but they were outside the scope of our analysis. However, whether any adjustment should be made for multiple testing remains a matter of debate (e.g. Sedgewick P. BMJ 2014;349:g6284), and as we have a) removed any mention of p-values/statistical significance from the manuscript text (in the previous response to review), b) now removed all p-values except p for trend (in line with Reviewer 2's suggestion), and c) the interpretation of the results is based on the direction of association and confidence intervals (not p-values), we do not feel that multiple testing is a major issue here.

Associations with trimester-specific exposures. The relevance of exposure during different trimesters is an important research topic within this field. These associations have been assessed and are presented in supplementary Table 9, but are no longer mentioned in the discussion. I think there needs to be at least some discussion of the relevance of the different trimester together with a summary of the evidence regarding timing from other studies. I agree with reviewer 3 that the paper by Rich et al. is one of the most persuasive papers on this topic and should be added to the discussion - irrespective of the 3rd trimester being most influential in the present study or not.

Authors' Response: We take on board the Reviewer's point, and have added the following text to the Discussion at line 517: "Our results did not give a clear indication as to which trimester could be most influential with respect to air pollution and fetal growth, and previous study findings have been inconsistent on this point. The most recent meta-analyses are suggestive overall of stronger associations for later trimesters between LBW/reduced birth weight and PM2.5(44) and PM10(45), but unclear for NO2.(1) One potential explanation for this is that earlier trimester exposures may be more prone to exposure measurement bias from residential mobility (in studies assigning exposure according to maternal residential address at birth), and thus attenuated towards the null. However, there are persuasive findings from a natural experiment of air pollution reductions during the 2008 Bejing Olympics, supporting the importance of the 3rd trimester exposures to air pollution in relation to term birth weight.(49-51) This is biologically plausible, as during the 3rd trimester the rate of fetal growth and weight gain increases dramatically and reaches its peak at about week 33.(50, 51)"

Reviewer: 2 - Michael Brauer

Comments:
The revised manuscript is much improved and presents a more balanced presentation of the study findings. This should be a valuable contribution to the literature. I agree with the authors approach to present results per IQR as this is most useful to compare impacts across the different exposures, including between air pollution and noise. I appreciated very much the inclusion of joint GAM models as these results are, in my view, among the most convincing. I'd suggest that their presentation in the manuscript could be more prominent.

Authors' Response: We thank the reviewer for these comments. In order to give more prominence to the results from the joint GAM models, as suggested by the reviewer, we have:
a) moved one set of joint GAM model figures (NO2 and night-time noise (Lnight)) into the main body of the paper (as Figure 5) so that these results are immediately available to the reader [for space reasons the remainder are presented in the Supplementary Material], and
b) reorganised the Results section between 249-313 to better integrate and balance the various analyses (noise as categorical, noise as continuous, and GAM models), and
c) amended the Discussion at line 360 to read as follows: "There was strong confounding of the relationship between road traffic noise and birth weight by primary traffic air pollutant co-exposures, and our results, particularly from GAM models, suggest little evidence for an independent exposure-response effect of traffic-related noise on birth weight outcomes."

Parity: The authors have now mentioned lack of information on parity as a limitation, but while the authors suggest this is most likely to be related to ethnicity or deprivation (which are accounted for), parity very easily could be spatially variable based on other factors such as maternal age and individual (not neighborhood) income. This could be added to the Discussion.

Authors' Response: We have added mention of maternal age and deprivation at both area- and individual-level at line 409 in relation to parity. We consider that we have already adjusted for deprivation at both area- and individual-level (the latter by proxy) as described at lines 399 to 407, but we have made this clearer in the text.

Air pollution model evaluation: The response to earlier comments and the revision regarding model evaluation should be further improved and authors may have misinterpreted the key point of the prior comment: it is not whether the model is or is not "valid" which is itself a question that cannot be answered directly (all are models and in fact they cannot be "validated" as that implies that truth is known, they can only be "evaluated" against some other measure) but rather whether the model performed better for some pollutants compared to others and if so how this might affect the comparisons between pollutant in the epidemiologic findings. If the model estimates of NO2 are less accurate than those for PM2.5 then that needs to be considered in the interpretation of epidemiologic findings. Within this context correlations are not the metric of interest but rather model RMSE, NMB or some related measure.

From the cited paper describing the air pollution model: "taking the root mean square error as an approximation of this uncertainty gives a value of 17.3 p.p.b. (28%) for NOX, 9.2 p.p.b. (36%) for NO2 and 1.6 μg/m3 (6%) for PM10" This suggests much better performance for PM10 than for NO/NOx. PM2.5 metrics are not reported in the cited publication but some metrics are in the cited internal report which also suggest higher NMB for NO/NOx compared to PM2.5. As such I reiterate the initial comment regarding the pollutant metrics and comparisons between them and suggest that conclusions regarding the impacts of one pollutant compared to another on the birth outcomes need to be addressed within this context and conclusions modified accordingly.

Authors' Response: We have changed lines 117 and 126 to refer to 'evaluation' rather than 'validation' in response to the Reviewer's comment. Thanks to the Reviewer for clarifying their earlier question - in response to this:

Re whether the model performed better for some pollutants compared to others:

The statistics picked out by the Reviewer above relate to model evaluation of annual mean predictions from reference 13. In this epidemiological study we are using monthly mean predictions, the equivalent evaluation statistics for which we have added to the manuscript at line 120 as follows: "Normalised mean bias (NMB) and root mean square error (RMSE) for modelled monthly predictions were slightly higher for NO_X (NMB = 11%, RMSE =13 μg/m^3 (22%)) and NO_2 (NMB = 11%, RMSE = 5.2 μg/m^3 (20%)) compared to $PM_{2.5}$ (NMB = 5%, RMSE =2.2 μg/m^3(14%)) and PM_{10} (NMB = 6%, RMSE =3.1 μg/m^3 (12%)), indicating that whilst all have a positive bias (NMB), $PM_{2.5}$ and PM_{10} are more accurately predicted than NO_2 and NO_X (RMSE)." The reviewer's point that NOX/NO2 predictions are more uncertain is correct, although not as marked as suggested from the annual data.

Re whether differences in model performance might affect the comparisons between pollutants in the epidemiologic findings:

We have added to the Limitations section of the Discussion the following text at line 390: "The air pollutant model predicted $PM_{2.5}$ and PM_{10} slightly more accurately than NO_2 and NO_X, but the model bias was in the same direction (over-prediction) for all these pollutants. Greater model prediction uncertainty for NO_2 and NO_X may result in effect estimates for NO_2 and NO_X being more conservative than those for $PM_{2.5}$ and PM_{10} and therefore may limit our ability to directly compare the magnitude of effect estimates for NO_2/NO_X vs. $PM_{2.5}$/PM_{10}."

However, given that the model bias is in the same direction for all of NO2/NOX/PM2.5/PM10 and the differences in model uncertainty are not particularly large, our interpretation in the Discussion that "…findings from two-air-pollutant models suggest that associations between term low birth weight and air pollutants emitted from vehicle exhausts may be driven by the fine particulate matter ($PM_{2.5\ \text{traffic-exhaust}}$) component rather than the gaseous NO_X component" remains reasonable, based on the consistency of direction of effect observed for PM2.5 traffic-exhaust alone (out of PM2.5 traffic-exhaust, NO2, NOX) after mutual adjustment. We have added text at line 440 re the consistency of direction of effect observed for PM2.5 traffic-exhaust to make clear this basis for our interpretation.

However we take the Reviewer's point (made below) that such statements may be too speculative for Conclusions sections, and so have removed them from the Conclusion/What This Study Adds section.

Spatial resolution of the noise model: To say that the model can be applied at a resolution of 0.1 m tells one nothing about the resolution of the model estimates themselves, unless the authors are actually arguing that the model can reliably detect differences in noise over a distance of 0.1m. Is the model actually developed at 0.1 m resolution (meaning that there are actually locations where estimates differ on that distance scale) or as I expect at some larger resolution (for example the paper describing the noise model indicates that some of the input data are at a resolution of 50m so is there some downscaling)? Most important would be some clear indication that the noise model estimates are indeed more highly resolved spatially (and this refers to the model itself, not its application) than the air pollution estimates as that would tend to support the findings by reducing the potential for exposure misclassification to favor the air pollution measures over noise in the joint models.

Authors' Response: Apologies for the confusion. In our previous response we meant to say that the spatial resolution of the noise model input data is 0.1m, however this is not the same as the resolution (or granularity) of the noise prediction. The noise prediction is calculated to 1 decimal place, i.e. 0.1 dbA, as per most noise models, e.g. SoundPLAN and commercial models. The 0.1 dBA resolution (or granularity) of the noise model is already stated at line 141. In terms of comparing noise and air pollution models: Both the noise and air pollution models have input data at the same spatial resolution, i.e. 0.1m. They both use OS MasterMap data which is the highest spatial resolution data available. The air pollutants are modelled at a spatial resolution of 20x20m. The mother's residential address birth is assigned the air pollutant concentrations for the nearest 20x20m grid point. The noise model predictions do not have a spatial resolution, per se, as they were modelled for specific address points for this study population, not on a regular grid basis. We have modified the text at line 367-373 to be more specific, and eliminate confusion, in this respect: "This study benefits from highly-spatially resolved air pollution modelling assigned at address-level, and noise levels estimated at address point. For noise particularly, this represents an advance on previous studies which have assigned noise exposure with lower spatial precision, e.g. postcodes,(8) 50m/250m buffers around maternal address,(7) or based on road proximity(29) and consequently reduces potential exposure misclassification, as noise levels may change dramatically over short distances (tens of metres)."

Abstract: Suggest removing from conclusions: "Our findings suggest that air pollution from road traffic in London, which is dominated by diesel exhaust emissions, is adversely affecting fetal growth. These results may be driven by the fine particulate matter (PM2.5 traffic-exhaust) component rather than the gaseous pollutants NO2 and NOX." While I do not disagree that diesel does dominate road traffic air pollution in London, this manuscript does not specifically address this point nor does the modeling as presented here indicate that diesel is more closely linked to PM2.5 (which the next sentence suggests is driving the impacts) than it is to NO2 or NOx. More generally both of these sentences are too speculative (see above comments regarding model evaluation) to be in the Conclusions of the abstract, although certainly it is reasonable to include them in the Discussion.

Authors' Response: We have modified the Abstract conclusion to remove mention of diesel and specific components of road traffic air pollution, as suggested by the Reviewer.

L238 (and elsewhere)– NO2 is generally not considered a primary pollutant

Authors' Response: The NO2 primary/not primary question depends on location. At the roadside NO2 is a combination of NO emitted locally and oxidised quickly by O3, NO2 emitted straight out of the exhausts of vehicles and background NO2 from other sources also oxidised by O3 – so at roadside it is a combination of primary (locally emitted) and secondary (oxidised by O3). In the context of this study which focuses on road traffic-related pollution in London we are using the term 'primary traffic-related pollutants' to convey a broad distinction between this group of air pollutants (NO_2, NO_X, $PM_{2.5\ traffic\text{-}exhaust}$ and $PM_{2.5\ traffic\text{-}non\text{-}exhaust}$) which are strongly influenced by local road traffic and much higher in London than in rural areas outside London, in contrast to $PM_{2.5}$ and PM_{10} which are much more highly influenced by regional particulate matter travelling hundreds of kilometres from source. Similar terminology and distinction has been used previously in the Health Effects Institute report on Traffic-Related Air Pollution (1). However, we have amended line 109 where these are first mentioned to read as follows: "NO_2, NO_X, $PM_{2.5\ traffic\text{-}exhaust}$ and $PM_{2.5\ traffic\text{-}non\text{-}exhaust}$ are primary traffic-related pollutants (i.e. locally emitted and/or rapidly formed near-source oxidation products)" to cover oxidation to NO2 of NO emitted from vehicle exhausts, for accuracy.

Table 2 still contains p-values for joint model results (OK to retain p for trend)

Authors' Response: Ok. We have removed the p-values (with exception of p for trend) from Table 2, and for consistency from the other tables containing regression model results.

L271 – this should only refer to Figure 2

Authors' Response: thank you for flagging this up. We have corrected this.

****

Additional changes: We noticed that the air pollutant results in Supplementary Table 7 were slightly incorrect, so we have corrected these – but there are only very minor changes to coefficient values, so it has no impact on interpretation/conclusions.

****

References

1. HEI Panel on the Health Effects of Traffic-Related Air Pollution. Traffic-Related Air Pollution: A Critical Review of the Literature on Emissions, Exposure, and Health Effects. Boston, MA: Health Effects Institute; 2010.