<u>RESPONSE TO COMMITTEE & REVIEWERS' COMMENTS</u>

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

We have addressed the comments from the Manuscript Committee and the Reviewers point by point below, and we feel that this has improved the manuscript. We provide clean and tracked changes versions of the revised manuscript. Our responses 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.

Report from The BMJ's manuscript committee meeting

Interesting RQ. 7-10 years old data, maybe difficult to get more recent data. Rescaled air pollution indices to interquartile range, becomes difficult to interpret effect sizes. Another source of confusion is role of confounding factors, e.g. smoking, no individual smoking data. The use of birth address - does this mean the address where the child was born, rather than the hospital? The effect size seems modest given the difference in birth weight versus use of interquartile range.

Assumption is most of time pre-delivery is present at home. The birth changes is dubious from a clinically meaningful viewpoint. Can they contextualise these changes or communicate the clinical significance better?

Authors' Response: We thank the committee for their thoughtful consideration of our manuscript. To cover the points above:

Re use of 7-10 years old data: Air pollution and noise exposures were modelled as part of a wider Traffic Pollution and Health in London project (https://www.kcl.ac.uk/lsm/research/divisions/aes/research/ERG/research-projects/traffic/index.aspx) which set out to investigate a range of health outcomes (cardiovascular disease, adverse birth outcomes, primary care data, incidence of heart attacks, hospital admissions and mortality). Exposures were modelled first, objectively, prior to examining health data. At the time exposure modelling was undertaken, the latest year of routine health data available was 2010 (for any of the outcomes of interest, including births). The years analysed reflect the long modelling timeframe and the lag in availability of routine health data. One of the problems was that the exposure modelling was done to end year of date of data available, i.e. 2010 – and takes so long to model on a 20m x 20m grid for London that it was not feasible to redo this. However, then we ran into the Care.data debacle and NHS Digital stopped giving out health data for two years and would not permit us link births registration data and hospital episode statistics (HES) data, which was our original plan. The HES data were required to provide gestational age and individual-level ethnicity information which is not available from birth registrations data. We then moved to using NN4B data, which was available from ONS and had ethnicity and gestational age, in place of HES data. By this time the original project was out of money and staff on the original project had moved on, so the whole project was delayed by several years.

Re rescaled air pollution indices to interquartile range, becomes difficult to interpret effect sizes: We have amended the paper to ensure that the IQR values are given in the footnotes for every figure/table as appropriate (this also addresses similar comment by Reviewer 1). However we also ran the analyses for pre-specified pollutant increments (e.g. 10 μg/m^3 for NO2, as described at line 191) to allow cross-trimester comparison and easy comparison with previous literature for specific

pollutants. These results are included in Supplementary Table 9, but we have made this clearer in the manuscript at line 327.

Re role of confounding factors, e.g. smoking, no individual smoking data: We address this below in response to your specific point raised.

Re birth address: we mean the maternal residential address at time of birth. We have clarified this throughout the manuscript.

Re Assumption is most of time pre-delivery is present at home: We acknowledge in the Discussion (line 432) that we could not account for exposures away from maternal residence (e.g. workplace, transport), and that this could contribute to exposure misclassification, as a potential limitation. There is a trade-off between choice of study designs: we conducted a very large registry-based study which has the advantages of complete case ascertainment, avoidance of selection bias, high statistical power, and relatively low cost, but relies on routinely collected data so information on workplace addresses/travel patterns are not available to incorporate into exposure assessment. It would only be possible to collect information on e.g. workplace addresses/travel patterns, in a much smaller scale study which involved direct participant recruitment and questionnaires (e.g. a cohort study), which would have the disadvantages of incomplete case ascertainment, likely low response rate and selection bias, lower statistical power, and relatively high cost. As Reviewer 1 points out "the large sample size and population-based study population are major strengths of the study".

Re "effect size seems modest given the difference in birth weight versus use of interquartile range" and "The birth [weight] changes is dubious from a clinically meaningful viewpoint. Can they contextualise these changes or communicate the clinical significance better?":
We take on board the Committee's point. There is meaningful value in examining continuous birth weight, however we acknowledge that for the general medical readership of the BMJ, the outcome term LBW has more immediate clinical significance, and therefore we have revised the paper to focus on the results for term LBW as the main outcome. This also reflects Reviewer 2's suggestion to focus on term LBW as the main outcome as it is well known as an important predictor of infant (and future adult) health. Additionally, to better convey the clinical significance of our findings, we now express the effect as a population attributable fraction at line 364 in the Results, and this is also highlighted in the Abstract at line 57 and Discussion at line 563. The population attributable fraction estimated for term low birth weight for exposure to $PM_{2.5}$ over the 25th percentile of the distribution (i.e. 13.8 µg/m^3) during pregnancy was 3% (0-7%). This 3% corresponds to 93 (0-216) cases of term low birth weight out of a total of 2950 cases per year on average in our London study population which are directly attributable to residential exposure during pregnancy to $PM_{2.5}$ >13.8 µg/m^3

Nonetheless we retain the results for continuous birth weight in the paper because they allow comparison with previous literature and they have meaningful value, particularly when the population contains different ethnic groups (for whom the 2500g cut off for LBW does not reflect constitutional differences between ethnicities, and therefore does not identify pathologically small babies equally in different ethnic groups). Most importantly, however, a small and clinically unimportant shift in a continuous distribution, e.g. blood pressure, birth weight or lung function, may nevertheless have an important effect at the population level on the proportion of individuals falling outside a clinical cut off. This has been demonstrated within our wider Traffic Pollution and Health in London project, i.e. exposure to traffic air pollution associated with a small mean reduction in lung function (FEV1), which increases the prevalence of children with abnormal lung function.(1) In this study we do observe a downward shift in the birth weight distribution, which suggests that air pollution has an impact across the whole population in terms of preventing fetuses reaching their full growth potential.

There are a number of specific issues relating to the statistical analysis and presentation of the data:

• **Are the confounders included in the statistical model sufficient? Only COA level data were available for deprivation and tobacco information. Is the lack of individual smoking data a serious problem?**
Authors' Response: As we describe in our manuscript, whilst we only have a direct measure of deprivation at COA level, we have also adjusted for individual-level birth registration type, which has been shown to relate to both qualifications and housing tenure at the individual-level (reference 37 Graham et al 2007). Graham et al (2) report that *"those sole registering births are typically lone parents (most cohabitants register jointly), facing high level of disadvantage (low qualifications, social tenancy, young age of mother). The low proportions receiving antenatal care, and having no-one else present at the birth of the child suggested some kind of exclusion or 'isolation'. There were strong links to social housing tenancy, though the causal ordering of birth registration and housing tenure (i.e. which comes first) is unclear."* We are confident therefore that we have adjusted sufficiently for deprivation/socio-economic status at individual-level, by proxy.

Area-level deprivation and individual-level deprivation represent different aspects/domains of deprivation, so it is important that we have adjusted for both.

Further, Graham et al 2007 report that birth registration type is also strongly associated with smoking at individual-level: *"Among those sole registering a birth, 63 per cent say they are smokers, compared with 45 per cent of unmarried jointly registering and just 16 per cent of married mothers."* As we have adjusted analyses for birth registration type at individual-level, we are confident that we have controlled at the individual-level for potential confounding by smoking, by proxy. Further, a confounder is by definition associated independently with both exposure and outcome – and the association between smoking and air pollution exposure is most likely through deprivation, which has been adjusted for at individual and area level. Tobacco expenditure has also previously been used as proxy for smoking in epidemiological research(3), and we have noted this in the manuscript at line 419.

We have expanded on this point in the Discussion at line 417-424.

• **The rescaling of air pollutant exposures to interquartile range increments makes it difficult for the general reader to interpret the effect sizes. Is this a conventional way of assessing air pollutant exposures?**
Authors' Response: to clarify, where multiple air pollutants are examined it is a conventional approach to rescale to the interquartile range, in order that effect estimates are calculated for comparable increases across the different pollutants (which may have very different absolute concentration ranges). We have explained this more clearly in the Methods section at line 193. As above, we also ran all the analyses for pre-specified pollutant increments (e.g. 10 μg/m^3 for NO2, as described at line 191) to allow cross-trimester comparison and easy comparison with previous literature for specific pollutants. These results are included in Supplementary Table 9, but we have made this clearer in the manuscript at line 327.

• **The actual effect sizes seem very small. For example, "IQR increases …were significantly associated with mean term birth weight reductions of 7 to 13g" (Abstract).**
Taking as an example ' PM2.5traffic exhaust' which has an IQR of 0.35 and mean 0.61 (Supplementary Table1), each IQR unit is 57% of the mean value. Hence, a greater than 50% change in the mean value is associated with birth weight reduction of 13g.

Authors' Response: As above, we have revised the paper to focus on term LBW as the main outcome as this better conveys clinical significance. In terms of effect size: IQR increases in primary traffic-related air pollutants, PM2.5 and PM10 were associated with 2-6% increased odds of term LBW. Taking NO2 as an example, the OR for term LBW was 1.03 (1.00, 1.06) for an IQR increase (8.6 μg/m^3) - this would translate to a ~20% change in the mean exposure value is associated with a 3% increase in risk of term LBW. For PM2.5, the OR for term LBW was 1.06 (1.01, 1. 12) for an IQR increase (2.2 μg/m^3) - this would translate to a ~15% change in the mean exposure value associated with a 6% increase in risk of term LBW. The effect sizes observed are consistent with previous research, e.g. the European-wide ESCAPE study[4] which observed an OR of 1.18 (1.06, 1.33) per 5 μg/m^3 increase in PM2.5 and ORs ranging from 1.04-1.09 (depending on analysis) per 10μg/m3 increase in NO2. Our results on the same scale (Supplementary Table 9) are OR of 1.15 (1.02, 1.30) per 5μg/m^3 increase in PM2.5 and OR of 1.04 (1.00, 1.07) per 10μg/m^3 increase in NO2.

- **There are a plethora of statistical significance tests. No acknowledgement of multiple comparisons is made.**

Authors' Response: We thank the committee for highlighting this omission. We have now acknowledged multiple comparisons in the manuscript at line 447.

Reviewer 1 - Ulrike Gehring

Recommendation:

Comments:
Summary
This study investigated the associations of air pollution and traffic noise with (low) birth weight and SGA in a more than 570,000 pregnant women living in Greater London and surrounding counties. The authors investigate potential confounding of associations with air pollution by traffic noise and vice versa.

The paper adds to the currently limited evidence on the association of traffic noise with pregnancy outcomes and the mutual confounding of air pollution and noise effects. It is we well written and the large sample size and population-based study population are major strengths of the study.

Authors' Response: We thank the reviewer for these comments.

Specific comments
Term births vs all birth: There seems to be some inconsistency. The abstract says that term (low) birth weight and SGA at term were the main outcome measures, whereas the method section states that associations were analyzed overall and in term births only without putting more emphasis pregnancy outcomes in term births. Figures and tables of associations of noise and air pollution with pregnancy outcomes in the main paper are provided for outcomes in term births only, which is consistent with the abstract, and associations in the overall population are summarized by a single sentence (page 12) without showing any of the results. I suggest to be consistent throughout the paper with regard to outcomes in term births being the main outcomes. Furthermore, associations for all birth should be presented in the supplemental material rather than stating "data not shown" as this increases transparency.
Authors' Response: We thank the reviewer for bringing this to our attention, and acknowledge that this is confusing. To clarify, outcomes in term births were our main outcomes and we have clarified this in the Methods section at line 203. We have reflected further upon this, and also the related point made by Reviewer 2, and have concluded that the all births analyses (conducted only for continuous birth weight and SGA) do not add value to the paper. The outcomes in term births are most important and meaningful to focus on from a clinical point of view – as they avoid mixing of poor growth and early birth (preterm) aetiological pathways towards low birth weight. To simplify and avoid confusion we have removed mention of the analysis of all births from the paper.

The paper would benefit from a map of the study area for readers that are not so familiar with the study area.
Authors' Response: we have added a map as Figure 1.

Air pollution exposures were available at 20m x 20m grid resolution, whereas noise exposures were assigned at address level. Potential exposure misclassification of especially air pollution from traffic, which is most consistently associated with birth weight in this study, may be a concern and should be discussed.
Authors' Response: Whilst there is still potential for exposure misclassification, the air pollution estimates modelled in this study are nonetheless highly spatially resolved – as the model produces a prediction on a regular grid of 20x20m – and they are assigned to address points geocoded to 0.1m resolution - thus reducing the potential for and the severity of any exposure misclassification, compared to studies which have less highly spatially resolved air pollution modelling (e.g. Olsson et al 2015(5), Hannam et al 2014(6)) and/or assign modelled estimates according to postcode centroids

[e.g. Stieb et al 2016(7)]. We have clarified at line 112 and 130 that air pollutants were modelled for specific points on a regular 20x20m grid (rather than an average per 20x20m grid square), and each address was assigned values of the nearest 20x20m point, as this may not have been clear previously.

There may be some exposure misclassification close to sources (where gradients of primary pollutants are steep). However, most people do not live within e.g. 10-30m of the centre of a main road so the impact on this study will be low. The percentage of maternal residences in our dataset within 10m of a major road (annual average daily traffic (AADT) > 10,000) was 0.07%, within 20m was 5% and within 30m was 11%.

We examined the relationship between living within 10, 20 and 30m of a major road and key individual-level variables (ethnicity, birth registration type, and maternal age). There was no association between living within 10m of a major road and ethnicity, birth registration type, and maternal age. The percentage of mothers (in our dataset) living within either 20m or 30m of a major road (as defined above) was slightly greater (by up to 3%) for non-White ethnicities (vs. White), unmarried mothers (vs. married), and younger (vs. older) maternal age groups. However, the percentage differences were very small (≤3%) between different categories, so there is no reason to assume that this would introduce serious bias.

Most importantly however, whilst there may be some exposure misclassification between the exposure at the actual address vs. the grid point estimate assigned, this should introduce no bias, because we have assigned the nearest 20x20m point. To introduce bias we would always have to choose the point closer to the road than the address location and this is unlikely.

As a result we do not think exposure misclassification is a major concern. We have discussed potential for exposure misclassification at line 396 onwards in the Discussion.

How was gestational age assessed? Based on ultrasound measures or based on last menstrual period? Please clarify.
Authors' Response: gestational age data were available from the NHS Numbers for Babies dataset (NN4B). No information about the method of gestational age assessment is recorded on NN4B records. The gestational age estimate is likely to be based on the more accurate and recent information from a mother's routine second trimester scan but a proportion may be based on the date of the last menstrual period(8). This has been clarified in the manuscript at line 93.

Parity may be a potential confounder of the observed association. If information on parity is available, it should be included. If not, this should be discussed as a potential limitation.
Authors' Response: Parity (but in particular nulliparity) is a risk factor for low birth weight. However, a confounder must by definition be associated independently with both exposure and outcome. An association between parity and exposure is most likely through ethnicity, or possibly deprivation, which are spatially patterned through London. However, associations between parity and ethnicity/deprivation (and thus exposure) are most likely only at higher order parity (i.e. parity ≥3) when cultural and economic factors are likely to become determinants of fertility. It is not, therefore, clear that parity would confound the observed associations – indeed it is difficult to predict how it would act. Information on parity was not available as part of this study, but as we have adjusted for both deprivation and ethnicity any potential confounding effect is likely to have been adjusted for. We have commented on this at line 424.

How did you use tobacco expenditure in your models? Continuous? Categorical? Please clarify.

Authors' Response: tobacco expenditure was included in the models as a continuous term. We have clarified this in the manuscript at line 206.

Statistical methods. Continuous vs categorical exposures: It is not clear why air pollution variables were used as continuous exposure variables and noise levels were categorized into 5 categories. Did you check the linearity of the exposure-response relationship and make the decision based on the outcome of these checks? If noise is linearly associated with the outcomes of interest, it is preferred to use it as a continuous variable, too, as categorization results in loss of power.

Authors' Response: There were two reasons for categorising noise. The first was that the noise distribution was highly skewed and could not be transformed to normality. Secondly, the noise model uses a constant for traffic on minor roads. This means that there is a lower level of variability in noise levels on some minor roads than in reality (the same could be said for some air pollutants). As a result, the noise model is likely to have over-estimated and under-estimated noise on some minor roads, but there is no geographical pattern (i.e. autocorrelation) in any bias due to this. As this could potentially hinder the detection of exposure–response relationships in linear analyses, particularly for night-time noise (for which noise levels are lower), the use of exposure categories may reduce exposure misclassification. We have added at comment to this effect in the Discussion at line 415.

Air pollutants, on the other hand, were normally distributed and to categorize these would have resulted in loss of statistical power, and lack of comparability with previous research. We do not think that categorizing both noise and air pollution would have made interpretation of the results any easier, and would simply have resulted in loss of power. However, we have checked the functional relationship between term birth weight and noise in response to the reviewer's suggestion (figures below for Lnight and Laeq,16hr) - there was some evidence of non-linearity, but it suggests our choice of categorical cut points are acceptable. By using categories we have not made any assumptions about linearity for noise in our analyses, although we have tested for a linear trend across categories, to give p-value for trend.
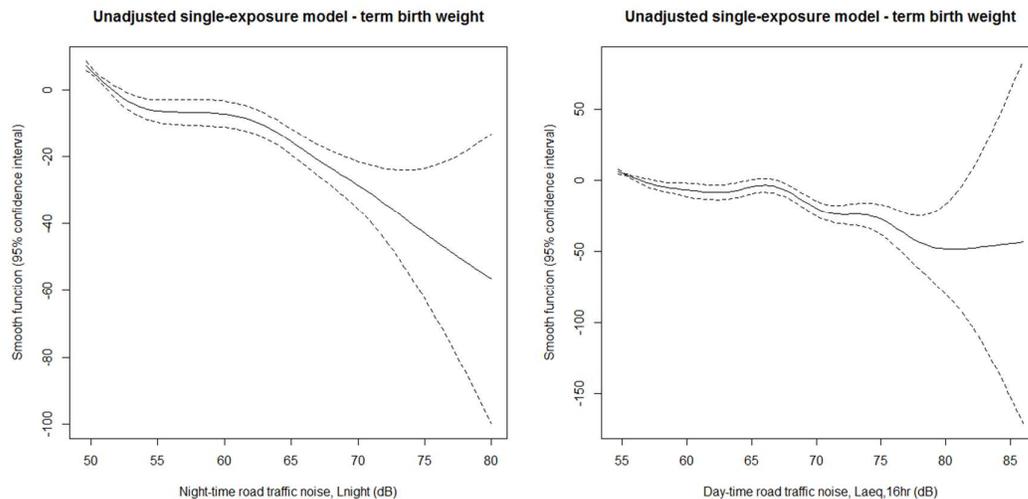
Since NO2 and NOx are perfectly correlated you may wish to consider leaving out one of the two.

Authors' Response: for transparency we have decided to keep both NO2 and NOx in the manuscript, firstly for transparency, and secondly to allow comparison with previous studies - some of which present NO2 but not NOx, and others NOx but not NO2.

Page 9, line 12. Add "singleton" after 572,910.
Authors' Response: amended as suggested at line 186.

Page 11, line 16. "Protective" suggests causality and should be replaced by a more neutral expression e.g. "An increase in birth weight was observed with increasing O3 exposure."
Authors' Response: amended as suggested at line 260.

Results, page 12, line 36ff. Add that confidence intervals for trimester-specific effects largely overlap.
Authors' Response: amended as suggested at line 326.

Results, description of Figure 3. Note that direction of noise effect changes upon adjustment for PM2.5 and PM10.
Authors' Response: We think the reviewer must be referring the models with day-time noise adjusted for PM2.5 traffic-exhaust and PM2.5 traffic non-exhaust (?), as we did not see any suggestion of this for the models adjusted for PM2.5 and PM10. However, as noted elsewhere we have added the full model results for noise (i.e. all exposure categories) to the figures rather than just the highest noise exposure category, as we realised it was causing confusion – and as a result it is clear that there is no overall pattern of change of direction across all noise categories when adjusting for PM2.5 traffic-exhaust and PM2.5 traffic non-exhaust, either for term LBW or term birth weight.

Multicollinearity checks are mentioned in the methods section, but outcomes of these checks are not provided in the results section. Please add.
Authors' Response: We have clarified the Methods section at line 213 on this point, and have added the results of the multicollinearity checks to the Results section at line 266. To clarify, we checked two air pollutant models for multicollinearity on a case by case basis first checking the VIF, and then the standard errors. Models with very high VIF for the exposure terms were excluded due to concerns re multicollinearity. Some VIF values were above 10 (a common rule of thumb), but borderline, so for these we then checked the standard errors. If the standard error more than doubled we excluded the model due to concerns re multicollinearity. However, for all two-air-pollutant models presented there was some increase in the standard errors for the exposure terms, which reflects the correlation structure between the air pollutants.

Results on exposure-ethnicity interaction should be shown (in the online supplement).
Authors' Response: We have added results of ethnic stratified analyses at Supplementary Table 10, and also amended the text in Results at lines 350-356 in line with this.   We have added a brief paragraph on this to the Discussion at lines 545-549.

Spontaneous vs medical intervention deliveries were not adjusted for or used to exclude or stratify analyses. To the extent that medical intervention is related to clinical practice variation which could vary spatially because of hospital catchment area or cultural factors, this is a serious limitation that could potentially confound the spatial component of exposure metrics. Medical intervention could influence all of the outcomes, directly (pre-term, very preterm) or indirectly (term birth weight via gestation length and SGA if medical intervention influences the clinical estimate of gestational age). See for example: Gagnon et al. Arch Gynecol Obstet. 2013 Apr;287(4):633-9. Hanley et al. Regional variation in the cesarean delivery and assisted vaginal delivery rates. Obstet Gynecol. 2010 Jun;115(6):1201-8. Janssen et al. Outcomes of planned hospital birth attended by midwives compared with physicians in British Columbia. Birth. 2007 Jun;34(2):140-7. `

Authors' Response: this is good point the reviewer raises, and we did consider the potential impact of this. Unfortunately, data on delivery type is not available as part of this study from birth registry data nor from the NN4B dataset. Although it would have been available if we had been permitted by NHS Digital to link hospital episode statistics (HES) data (see earlier comment re issues with NHS Digital data provision on page 1). However, all our epidemiological models were mixed models which included a random effect for small area (MSOA – which on average covers a population of 8000 people) specifically to account for underlying spatial patterns in the data, so in our opinion we should have adjusted for this sufficiently. Nonetheless, we have commented upon this as a possible limitation of the study in the Discussion at line 439. Compared with the Canadian studies noted by the reviewer, the UK NHS medical system reflects a relatively small geographical area – and a much more uniform system than in Canada.

Table 1. Tobacco expenditure is missing here. Please add.
Authors' Response: we have added this to Table 1.

Table 2. This is probably minor given the high correlation between daytime and nighttime noise, but are air pollution effect estimates adjusted for daytime or nighttime noise? I suggest to have 2 separate tables one for daytime noise and one for nighttime noise, one of them could be moved to the supplement.
Authors' Response: We have revised Table 2 to show air pollution effect estimates adjusted for both night-time and day-time noise. There are no differences between the air pollution estimates adjusted for night-time noise vs. adjusted for day-time noise.

Add crude effect estimates to Figure 1.
Authors' Response: we have added these to what was Figure 1 (but is now Supplementary Figure 3).

Figures 1-3. Add a footnote with the exact increments for this analysis.
Authors' Response: we have added footnotes as suggested to all figures.

Supplementary Tables 2 and 3. Clarify increments for gestational age (is it 1 week?) and tobacco expenditure. Clarify that maternal age is in years.
Authors' Response: The increments are per 1 completed week for gestational age, and per £1 increment for tobacco expenditure. We have clarified this in Supplementary Tables 2 and 3.

Table 2 says "gestational age", table 3 says "gestational age (completed weeks)". What is correct? Revise accordingly.
Authors' Response: these should all say completed weeks, we have amended this. Thanks for highlighting.

Table 2, supplementary table 4-6. Replace "per IQR" by the exact increments used.
Authors' Response: we have added the IQR increments to the footnotes for all tables showing analysis of air pollutants on the IQR scale, but we feel it is important that the Tables clearly state 'per IQR' so that the reader knows that the increments are comparable for each of the air pollutants.

Reviewer 2 - Jo Kay Ghosh

Recommendation:

Comments:
The manuscript presents results from a large, registry-based study of traffic-related air pollution and traffic-related noise pollution and effects on several measures of birth weight (term birth weight, term low birth weight, term SGA). The study was conducted in London metropolitan area, using birth registry data for outcome evaluation and a dispersion modeling approach for exposure estimation.

The question of joint effects of noise and air pollution impacts of birth weight is an area that is not well studied, as the authors correctly identified, so this manuscript has the potential to add information to this topic. However, there are some key areas of the manuscripts that should be addressed. First, there is too much emphasis on statistical significance of the results. Instead, the focus should be on reporting on the estimated effects and associated confidence intervals. Second, it is not clear why the authors focused on these measures of birth weight. Biologically, is there evidence that a 7-13g reduction in birth weight is important to health (for birth weights across the distribution)? There is substantial evidence, however, that term LBW is an important predictor of infant (and future adult) health, so this would be my recommended outcome measure to use. In the Results section, the authors also mention doing an analysis for birth weight and SGA (not term), without explanation of the relevance of this analysis. Third, there are a number of studies that also examine the role of heat/temperature on birth outcomes, so that may be an important factor to consider.
Authors' Response: We thank the reviewer for these thoughtful comments. To cover the points above:

Re too much emphasis on statistical significance of the results. Instead, the focus should be on reporting on the estimated effects and associated confidence intervals. We take on board this point, and have removed mention of "significant" findings and p-values from the text of the manuscript (except for reporting interaction terms), and shifted the focus to estimated effects/confidence intervals as suggested by the Reviewer.

Re birth weight outcome measures: We take on board these points made by the Reviewer. As mentioned earlier, we have changed the emphasis of the paper to focus on the results for term LBW, to better convey clinical significance. However, we still retain the results for continuous birth weight within the paper, for reasons described earlier. Re analyses on term vs all births – again, as mentioned earlier, we have concluded that the all births analyses do not add value to the paper, and therefore we have removed any mention of them.

Re the role of heat/temperature on birth outcomes: There is some evidence to suggest that extremes of ambient temperature may be associated with adverse birth outcomes: e.g. preterm birth/early delivery(9-11) and low birth weight(12, 13). Meteorological conditions, including ambient temperature, are related to air pollution levels. By adjusting for season we did adjust for general seasonal variation in average temperatures, but we could not adjust for exposure to extreme ambient temperatures as we did not have data on these. We have commented on this at line 427 in the Discussion.

Specific comments:
Methods - it is not clear why the study area was strictly limited to the area within the M25 boundary. Excluding the births in the areas overlapping the M25 would have excluded populations

that were most highly exposed to pollution from this source. This would be an important group to include, if data are available.

Authors' Response: The study boundary was the M25, because we did not have traffic information beyond the M25 and therefore no air pollution or noise estimates beyond this, we have clarified this at line 86 in the paper. We linked a range of area-level potential confounding factors, some at MSOA-level and some at COA-level, to birth records. We needed to further restrict the study area to all MSOAs <u>fully</u> within the M25, because some of the MSOAs are very large and extend beyond the M25 by quite a large area, and this would mean we include confounder information for a population which might be quite different in characteristics from our study population (urban vs rural; white vs ethnically diverse etc.). our map which we have now added as Figure 1 shows, the study area does extend right up to the M25 in some places, so the study population will include population who are most highly exposed from the M25.

Methods - Is the population-based metric of tobacco expenditures a good proxy for maternal smoking, or perhaps second-hand smoke exposure? Providing a reference that validates this as a proxy measure would be helpful.

Authors' Response: Tobacco expenditure has previously been used as proxy for smoking in epidemiological research(3), and we have noted this in the manuscript. However we do not have information on how good a proxy it is for maternal smoking or second-hand smoke exposure. Nonetheless, as described above, we have also adjusted for birth registration type (individual-level variable) which is strongly associated with maternal smoking at the individual-level, and therefore we are confident that we have controlled sufficiently for maternal smoking, by proxy.

Methods - p. 9 - For the births missing noise exposure and/or ethnicity (since these categories included 30,000+ records), it would be better to include them in the analysis and use imputation methods to address missing data.

Authors' Response: whilst the numbers missing for noise exposure or ethnicity may sound large in absolute terms, it is important to remember that this study has a very large sample size, and as such the missing data for noise/ethnicity represents only about 5-7%. Imputation methods are not a perfect solution to address missing data as they introduce error, and such methods are only necessary if the proportion missing data is very large, or, the sample size is small – and neither of those apply in this case. We have checked the data and there is no reason to think that they are missing not at random (MNAR), so excluding those missing data should not bias the results after adjustment.

Methods - p.9 - How were the cut points for noise determined? Are these biologically relevant categories?

Authors' Response: we categorised noise by 5 dB increments, in line with previous analyses conducted in the wider Traffic study(14), and in London(15). There is insufficient evidence to date to be able to define biologically relevant categories in relation to birth outcomes. More generally, however, adverse health effects are observed in the 40-55 dB range for $L_{night,outside}$, and above 55dB adverse health effects occur frequently(16), so we ensured we included 55dB as one of the cut points.

Methods - p.9 - Were the adjustment factors in the models determined a priori, and why were these selected?

Authors' Response: all models were adjusted *a priori* for maternal age; birth registration type; Carstairs deprivation quintile; and tobacco expenditure. Continuous birth weight and LBW models were also adjusted *a priori* for sex, gestational age (linear and quadratic terms), and baby's ethnicity. These decisions were made based on prior knowledge/existing literature which suggested that they could potentially confound the relationship between exposure and outcome. SGA models were not

adjusted for sex, gestational age or baby's ethnicity because those factors are already accounted for in the calculation of SGA. Adjustment of noise for air pollution, vice versa air pollution for noise, was also determined *a priori* as it was integral to the study aim to examine joint air pollution and noise exposures. Season of birth, year of birth and a random effect for small area (random intercept for MSOA) were identified, *a priori*, as potential factors to consider adjusting for, but the decision as to their inclusion was based on testing how influential these factors were in the model. We have updated the Methods text, at line 209, to clarify how model adjustment decisions were made.

Discussion - p. 15 - Diesel PM includes a number of air toxics, so EC may not be the biologically active component causing the health outcomes of interest. Rather, EC is typically used as an indicator of diesel exhaust exposures, i.e. the exposure to the mixture of toxics contained in diesel exhaust.
Authors' Response: we take on board the reviewer's point here and have deleted the sentence referred to.

General - The manuscript contains awkward phrasing in some areas (e.g. "Exposure is complex, close to roads and individual would be exposed to more primary exhaust and non-exhaust... particles...", "However, traffic also produces noise, which has been associated with e.g. hypertension and cardiovascular disease") and some grammatical issues that are distracting to the reader.
Authors' Response: we have amended the phrasing highlighted by the reviewer at lines 66 and 72.

Reviewer 3 - Michael Brauer

Recommendation:

Comments:
This is a well-written manuscript, describing a well-conducted and comprehensive study to assess the relationship between near-road exposures to traffic-related air pollution and the potential for independent effects and/or confounding from traffic noise. This study adds to a growing and maturing literature suggesting causal relationships between air pollution and reduced term birthweight.
Authors' Response: We thank the reviewer for these comments.

Air pollution exposures were estimated from a dispersion model. This model is likely more reliable for primary traffic vs related secondary emissions and perhaps also more accurate for gaseous vs particulate pollutants. Some of the authors have published land use regression models for multiple pollutants in London and it would be helpful to evaluate if the observed associations are also evident with these other exposure estimates (which in other locations have proven to more accurately reflect measured concentrations than dispersion model estimates).

In addition there really needs to be some mention of model evaluation with measurements – the only reference is to the model website, not to any peer-reviewed publications, and no mention at all of evaluation. More importantly, what is the accuracy of the model estimates and are they equally accurate for the different pollutants that were evaluated? If not then this must be considered in the interpretation of the epidemiologic findings when comparing the different pollutant metrics. Treating all pollutant measures as equally accurate masks what could be important (and non-random) exposure misclassification.

Authors' Response:
Comparison with land use regression models: A comparison between LUR models and the KCLurban dispersion model is outside the scope of this analysis - we are not evaluating the model because it has already been evaluated and shown to be valid (see below). There is a separate project currently ongoing "Comparative evaluation of Spatio-Temporal Exposure Assessment Methods for estimating the health effects of air pollution (STEAM)" led by Klea Katsouyanni at Kings College London, which will compare LUR with dispersion modelling outputs from the Traffic Pollution and Health in London project.

Model evaluation and accuracy: To clarify, we did reference a peer-reviewed publication (reference 13 in the manuscript) for the air pollution model which includes model evaluation with measurements, and the full model evaluation (Reference 15 in the manuscript) which is a detailed supplementary file available open access via the Traffic Pollution and Health in London project website. However, we did not explicitly mention model evaluation in the manuscript, and we have now rectified this, starting at line 121. The model performed well when validated against measurements, with very low normalised mean bias and high spearman correlation coefficients (r) between observed vs modelled monthly concentrations: r>0.91 for NOX, PM10 and PM2.5, r>0.83 for NO2 and r>0.9 for O3 at both roadside and background locations. The model doesn't always predict NO2 and PM2.5 equally well, not least because they are evaluated at different numbers of sites.

The air pollution model has already been used for epidemiological analyses for a range of other health outcomes, and we have added reference to those 5 peer-reviewed publications in the manuscript (References 16-20 in the manuscript).

For balance, we have also added a sentence re noise model evaluation at line 147. Model validation studies conducted in two UK cities demonstrated high Spearman's correlation (0.90) between measured and modelled noise levels, indicating good model performance.

What was the spatial resolution of the noise model? The manuscript suggest reduced exposure misclassification compared to other studies due to the use of addresses. This is valid but perhaps a bit oversold - the air pollution model has resolution of 20m so not fully resolved to each address and no indication is given on the resolution of the noise model, except to indicate that estimates were generated for each address.
Authors' Response: To clarify, addresses were geocoded to a spatial resolution of 0.1m. Because the address point XY coordinates fall on the geometric centroid of the dwelling, the address points were universally moved to 1m from the façade of the dwelling at the front of the property. The spatial resolution of those address points remained at 0.1m. Noise was modelled for those exact address points 1m in front of the façade, and therefore the spatial resolution of the noise modelling is 0.1m. We have clarified the spatial resolution of the address geocoding (and thus the noise modelling) at line 149. Noise is therefore slightly better spatially resolved than air pollution.

The air pollution metrics were analyzed as continuous exposures whereas the noise metrics are analyzed categorically. By treating the exposures differently it makes it rather difficult to accurately determine the impact of adjusting for one exposure vs the other – I'd suggest the authors pay more attention to this and would be well-advised to evaluate non-linear models for both exposures (including in joint exposure models) so one can address the shape of the relationships and evaluate whether the current categorization of noise exposures was appropriate. This issue could have bearing on the interpretation of the results.
Authors' Response: As noted above we have examined the functional relationship between term birth weight and noise, and we think the categorization of noise exposure was appropriate. Further, by using categories we are not making any assumptions about linearity for noise. However, in response to the reviewer's comments we have run adjusted generalized additive models (GAMs) for adjusted single and joint exposure models to examine the shape of the relationships, and to see if our findings hold when we make no assumptions about the linearity of the exposure-response functions for either air pollution or noise. The results were as follows:

In adjusted single-exposure GAM models, term birth weight decreased with increasing exposure to air pollutants (except O3) or noise, in a largely linear fashion. This corresponds to our findings from linear regression models.

In adjusted GAM models for joint exposure models, air pollution effects were robust to adjustment for noise, but once adjusted for air pollution there appears to be no relationship between noise and term birth weight. This corresponds to our findings from linear regression models.

We are therefore satisfied that our conclusions based on analysis of continuous air pollutants and categorical noise metrics are robust, but we thank the reviewer for making this suggestion which has helped to strengthen the conclusion.

We have added to the Methods at line 215 and the Results at line 312, and the plots from the adjusted GAM models for joint exposure models in Supplementary Appendix 1 to cover these additional analyses.

The general conclusion of noise associations representing confounding is not entirely clear – the results (Figure 2) suggests attenuation of noise effects when air pollution included in models but there still does appear to be some suggestion of association for nighttime noise (especially in

models including NO2/NOX). This result should get a bit more prominence in the Abstract and Discussion, or minimally some more discussion of why the authors conclude that all noise effects seem to be confounded. The logic as presented would seem to be: primary exhaust PM2.5 drives the air pollution associations (in the air pollution only models) so consequently the NO2 associations in the joint (with noise) models are not given as much attention compared to the models with primary exhaust PM2.5. This is not really valid either statistically or from a logic perspective – as it is more likely that primary PM2.5 and NO2 reflect different aspects of the traffic pollutant mixture which may have different relationships with traffic noise.

Authors' Response: There is a difference in the correlation between noise and NO2/NOx vs. noise and primary exhaust PM2.5, and the different joint exposure models could represent adjustment of the noise associations for different aspects of the primary road traffic air pollution mixture. The reviewer therefore has a fair point re the logic of emphasising the noise models adjusted for primary PM2.5 exhaust, so we have modified the description of the results at line 299-310, and in the Discussion at line 480-486, to be more balanced with respect to the models adjusted for NO2/NOX.

The Reviewer's comments are helpful in making us realise that presenting only the estimate for the High noise category in this figure may perhaps be misleading. We have therefore added the full models results for noise (i.e. all exposure categories) to the figures [Figures 2, 4, Supplementary Figures 2, 3] rather than just the highest noise exposure category. For transparency, we also provide the results in tabular form in Supplementary Table 8, so that the reader has the exact effect estimates available. With the full noise model in the figures, we think it becomes clearer that there is little overall to suggest an exposure-response relationship of term birth weight with noise after adjustment for any of the primary road-traffic-related air pollutants – although we acknowledge there remains a suggestion of reduced term birth weight in the high night-time noise category in models adjusted for NO2 or NOX (Supplementary Figure 2, which was previously Figure 2 referred to by the Reviewer). We have added the corresponding figures for term LBW (Figures 2 and 3), and these do not suggest an association between term LBW and noise, after adjustment for NO2 or NOX.

Our findings from non-linear GAM models, which we have run on the suggestion of the reviewer, and which are added to this paper at Supplementary Appendix 1, support our conclusion that there after adjustment for any of the primary traffic-related pollutants (NO2, NOX, PM2.5 exhaust, PM2.5 non-exhaust) no association remains with night-time noise. In GAM models adjusted for NO2 or NOX, there is no suggestion of a threshold effect for noise.

However, we modified the Abstract as follows to give a more nuanced presentation and interpretation of our results:

- at line 48 we now say that noise associations were "strongly attenuated" rather than "attenuated to null" when adjusted for primary traffic air pollutants.
- at line 53 we now say "Our results suggest little evidence for an independent exposure-response effect of traffic-related noise on birth weight outcomes." rather than "An apparent relationship with noise appears to reflect confounding by primary traffic air pollutant co-exposures."

We have amended the Conclusion to read "Our results suggest little evidence for an independent exposure-response effect of traffic-related noise on birth weight" rather than "Our results, from by far the largest study to date, found no independent effect of traffic-related noise on birth weight, and suggest any apparent relationship reflects traffic air pollutant co-exposures." The What This Study Adds box is similarly amended.

As there were also associations present with non-exhaust air pollution it would seem that the manuscript should not be as forthright regarding the observed associations being related to diesel emissions and traffic exhaust (vs non-exhaust) findings (Abstract, What this Study Adds, Discussion).

Authors' Response: We take on board the reviewer's point and have balanced out the discussion regarding conclusions that can be drawn re exhaust vs non-exhaust air pollution at line 468 in the Discussion, as follows: "Our findings suggesting that associations between low birth weight and air pollutants emitted from vehicle exhausts, may be driven by the fine particulate matter (PM2.5 traffic-exhaust) component rather than the gaseous pollutants NO2 and NOX is an important and novel contribution to scientific knowledge. Our study also demonstrates associations between LBW and fine particulate matter from road traffic which is not emitted from the vehicle exhaust, i.e. brake/tyre wear particles and vehicle-induced resuspension of road dust. However, due to multi-collinearity in models containing both PM2.5 traffic-exhaust and PM2.5 traffic-non-exhaust, we could not disentangle potential effects of traffic-related exhaust and non-exhaust related PM2.5. The magnitude of association with PM2.5 traffic-exhaust was consistently stronger than with PM2.5 traffic-non-exhaust, and this could reflect differing chemical constituents (and thus toxicity) of the PM2.5 mixture from different sources." We have amended the Abstract (lines 53-57) and What this Study Adds box in line with this.

The authors should tone down the mentions of this being the largest such study – large doesn't mean better and power is not really a concern for outcomes such as term birthweight or SGA. Both the Canadian and Danish studies, while smaller, were also population-based/included the entire population. It is OK to mention once that this is largest study but four times in the manuscript is a bit much.

Authors' Response: We have amended the manuscript to limit mention of this to the start of the Discussion.

There are multiple mentions in the manuscript of "significant" findings and also presents p-values. To conform to current best practices in epidemiology and with the statement of the American Statistical Association on statistical significance and p-values (https://www.amstat.org/asa/files/pdfs/P-ValueStatement.pdf) I'd suggest removing mention of significance and p-values throughout the manuscript. The authors have appropriately presented effect estimates and confidence intervals to allow for proper interpretation by readers.

Authors' Response: We have removed mention of "significant" findings and p-values from the text of the manuscript (except for reporting interaction terms) and shifted the focus to estimated effects/confidence intervals, as suggested by the Reviewer. We do think the p-value provides some valuable information for the reader regarding the statistical incompatibility of the data with the null hypothesis, and thus we have retained p-values in the results tables – this is in line with the "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.

There seems to be a missed opportunity in the study to separately examine the impact of aircraft noise measures such as those used in prior study conducted by several of the same authors (Hansell AL, Blangiardo M, Fortunato L, Floud S, de Hoogh K, Fecht D, Ghosh RE, Laszlo HE, Pearson C, Beale L, Beevers S, Gulliver J, Best N, Richardson S, Elliott P.Aircraft noise and cardiovascular disease near Heathrow airport in London: small area study. BMJ. 2013 Oct 8;347:f5432. doi: 10.1136/bmj.f5432.) Use of these measurements (and related modelled contours) would add much to the current analysis in their ability to isolate noise from aircraft from the somewhat correlated traffic noise and traffic air pollution. The authors have included a sensitivity analysis

where those with high exposures to aircraft noise were excluded (and not surprisingly found little impact on results given that this must represent a minority of the population) but much more useful would be a sensitivity analysis in which the subset of those exposed to high aircraft noise were specifically evaluated.

Authors' Response: We agree with the Reviewer that the impact of aircraft noise on birth weight would be an interesting question to investigate separately. In fact, the authors mentioned by the Reviewer are currently investigating that specific research question separately as a PhD studentship.

It was not, however, our research question in this particular study. We had only a binary indicator of exposure to aircraft noise (≤50dB vs. >50dB), for the purposes of sensitivity analysis. We conducted a sensitivity analysis to remove the influence of high aircraft (and rail) noise, and thus give a 'cleaner' sub-group who were not exposed to high levels of noise from other transport sources (aircraft and/or rail) for analysis. Exclusion is the correct way to do the sensitivity analysis when the specific research question under investigation is to examine road traffic noise.

In response to the reviewer's comment that "a sensitivity analysis where those with high exposures to aircraft noise were excluded (and not surprisingly found little impact on results given that this must represent a minority of the population)" – we clarify that the proportion of the population exposed to high aircraft noise was not a minority, but a considerable proportion: in total, the sensitivity analysis excluded N=113257 (21%) who were exposed to aircraft and/or rail noise >50dB (with 14% exposed to aircraft noise > 50dB, and 8% to rail noise > 50dB). It is therefore notable that after excluding 21% of the population, the findings remain essentially unchanged.

The authors may want to add to the Discussion the rather persuasive findings from a natural experiment which supports the suggestion of the importance of 3rd trimester exposures, as reported in the manuscript (Rich DQ, Liu K, Zhang J, Thurston SW, Stevens TP, Pan Y, Kane C, Weinberger B, Ohman-Strickland P, Woodruff TJ, Duan X, Assibey-Mensah V, Zhang J. Differences in Birth Weight Associated with the 2008 Beijing Olympics Air Pollution Reduction: Results from a Natural Experiment. Environ Health Perspect. 2015 Sep;123(9):880-7. doi: 10.1289/ehp.1408795)
Authors' Response: We thank the reviewer for this suggestion – however, as we have now added the trimester specific analysis for LBW and SGA, it is less clear which trimester is most influential. Therefore we have not added this reference.

Reviewer 4 – Charlotte Clark

Recommendation:

Comments:
This is a well carried out study with good data linkage and coverage. This is a novel study, one of the largest to date on noise exposure and birth outcomes. The analyses are well executed, and a lot of data is presented. I feel that some of the interesting discussion has been constrained by the word-length of articles in the BMJ.
I would recommend publication of this paper in the BMJ after some minor changes to the paper.

Abstract
Consider revising the first few lines of the results section. It is very hard-going for someone coming anew to the paper, probably given the level of detail relating to the various metrics..which whilst I appreciate are important, I just don't think it reads very easily.
Authors' Response: we thank the reviewer for this feedback, and have revised as suggested.

Conclusions: don't the findings that the relationship with noise apparently appears to reflect confounding by traffic air pollutant exposures conflict with the findings of recent reviews of combined exposures for other outcomes (Tétreault et al., 2013; Stansfeld, 2015). Why might this be? Is it because of something specific to the outcome here as the other papers tend to deal with cardiovascular outcomes? The reviews are limited in number of studies and size of studies. These issues should be elaborated on in the discussion.
Authors' Response: This is a very interesting point, which we now discuss at lines 526-543. In the broader context, our findings do contrast with reviews of joint air pollution and noise studies which suggest independent effects of road traffic noise on other health outcomes, e.g. cardiovascular outcomes, after adjustment for air pollution.(17, 18) This could reflect different biological pathways between noise and fetal growth vs. other health outcomes at different stages in life. The fetus has no direct exposure to the environment, but exposure is mediated via the mother and placenta, and this may modify effects. Alternatively, threshold effects may be relevant for exposure to noise, and the threshold could vary between health outcomes, possibly being higher for effects on birth weight vs. e.g. cardiovascular outcomes. Equally, it might reflect differences between studies in the ability to control for confounding by air pollution from road-traffic specifically. We did note that associations between noise and birth weight were more strongly attenuated by adjustment for primary road traffic related air pollutants (NO2, NOX, PM2.5 traffic-exhaust, PM2.5 traffic-non-exhaust) compared to background air pollutants (PM2.5, PM10). This suggests that adjusting for the background pollutants may not fully adjust for the confounding effects of air pollution co-exposures directly from road traffic, in our study. With respect to cardiovascular outcomes, it has been noted that "more studies using air pollution indicators specific to road traffic are needed to properly assess if road noise and pollutant effects on CV outcomes are subjected to the confounding effect of one another."(18)

Introduction
Could something brief about potential mechanisms for how air pollution and noise might biologically influence birthweight be added. I appreciate BMJ intros are short but this would benefit the reader for whom it may not be immediately obvious how noise influences birthweight especially.
Authors' Response: As suggested, we have added a sentence re potential noise mechanisms at line 75.

Method

Section on noise exposures – I wasn't clear how moving house was taken into account in these estimates. This is later addressed in the discussion but should be clearly stated here. It is also worth checking the terminology in this section as I'm not sure that 'receptor placement' would be understood by a general audience.

Authors' Response: We have clarified that we did not have information on whether a mother moved address during pregnancy in the Methods section at line 99. We have also clarified 'receptor placement' at line 156.

Section on outcomes – I think these are the standard WHO definitions for birthweight but worth referencing. The last part of this section, where the N's are given for missing data – could % also be provided to help the reader understand more easily the extent of lost data.

Authors' Response: we have added a reference to the WHO definition for low birth weight at line 163. We have added % for missing data as suggested, but in order to do so we had to change the exclusion Ns to totals, rather than Ns for sequential exclusion, in order for this to make sense. So the exclusion Ns are no longer mutually exclusive.

Statistical methods – the categorisation of the Lnight – the lowest group is <50 dB. This is above the WHO NNG – was there not power to have a group that represented the NNG level of 40dB? I'm assuming not but could be worth discussing NNG in more detail.

Authors' Response: To clarify, the minimum estimated value of Lnight for our study population was 49.6 dB, as shown in Supplementary Table 1. We did not have estimates lower than this value because of the 'floor' imposed in the model. The paper we have referenced(19) which describes the TRANEX noise model explains this: "The CoRTN method implemented in TRANEX predicts a minimum night-time LAeq,1hr of 38.0 dB for a single minor road. The minimum modelled value of night-time LAeq,1hr in London was 42.4 dB because we combine the contribution of all sources up to 0.5 km or 1 km." The model floor means that the minimum modelled value of night-time noise from traffic in London was 42.4dB, which is higher than the recommended upper limit of exposure of total noise of 40dB proposed by the Night Noise Guidelines for Europe. It is possible that we did not have a sufficiently low noise exposure reference group, to detect small associations between noise and birth weight, above the guideline level. We have commented on this at line 519.

Results

Are these levels of LBW and SGA what we would expect compared with national norms?

Authors' Response: According to the Office for National Statistics in 2008 (our midpoint year), the % of live births <2500g was 7.2 % across England and Wales, and 7.6% in London (https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/livebirths/datasets/birthsbyareaofusualresidenceofmotheruk). There is variation in the UK by region. The levels of LBW in our study population (after data cleaning of birth weight and prior to exclusion of multiple births) were very similar (7.5%) to the ONS national and London statistics. The 5.7% LBW in Table 1 and stated in the results is lower that ONS statistics, largely due to the exclusion of multiple births (which remain in the ONS calculations). So once that is taken into consideration, yes the levels of LBW are what we would expect compared to national norms.

There are no national data on SGA for comparison. However because SGA is defined as lower than the 10th centile for a given gestational age, the prevalence will always be in the region of 9-10%, as it is in our study population.

I wasn't very sure why the sensitivity analyses for those exposed to rail/air noise above 50dB included. Perhaps make the reasoning stronger here. The level of noise selected here at the lower

end seems quite a low exposure and it seems a little crude to categorise everyone together who is 50dB or more.

Authors' Response: As our research objective was specifically to examine air pollution and noise from road traffic, we conducted this sensitivity analysis to remove the influence of high aircraft and rail noise, and thus give a 'cleaner' sub-group who were not exposed to high levels of noise from other transport sources (aircraft and/or rail) for analysis. We have clarified the reasoning at line 221.

It is crude, we agree. However we did not have available to us any quantitative information on rail/aircraft noise below 50dB, as this is not modelled by the data providers. Ideally we would have excluded above the average of the true distribution, but we could not calculate the true average due to the artificial floor (minimum) on the data available to us.

The absolute noise level of the cut off may seem low, but the proportion exposed to >50dB rail/aircraft noise was 21% (with 14% exposed to aircraft noise > 50dB, and 8% to rail noise > 50dB), so the cut off does actually represent the highest deciles of the distribution of exposure to aircraft and rail noise in the study population.

Similarly, whilst I appreciate that birthweight is associated with ethnicity, as noise and air pollution can be, I wasn't sure why we might expect the association between these exposures and birthweight to differ across ethnicities.

Authors' Response: It is not uncommon to see effect modification by ethnicity in epidemiological studies. Effect modification of the relationship between air pollution and reduced birth weight has been observed previously, although the results are inconsistent: stronger associations in Black or Hispanic groups vs. other ethnicities (20-23), stronger associations in the White group vs. other ethnicities (24), or variation in effect modification by ethnicity according to pollutant (25) have all been observed. Diverse ethnicity, spatially patterned, is a key characteristic of London. We adjust models for it to adjust for any confounding by ethnicity, which could be due to spatial patterning/socio-economic status/cultural factors. However, by stratifying we also address questions about effect modification, i.e. does susceptibility to, or response to, environmental insults differ according to ethnicity – which could reflect environmental inequality or differing biological susceptibility.

Discussion

I thought that noise got little discussion time in the discussion section. The key questions the paper raised for me that I didn't feel were well addressed were:

why would air pollution be more important for birthweight than noise?

why might air pollution only show an effect for birthweight, when for other outcomes such as cardiovascular health both exposures seem equally important and independent?

Is a -7 to -13g difference in birthweight important – I imagine it is but this isn't addressed.

Authors' Response:

As to, "**why would air pollution be more important for birthweight than noise?**": In terms of public health, air pollution and noise may both be important, because they are both widespread environmental exposures in urban environments. Whether they turn out to have equal impact on public health, depends upon the association they each have with birth weight. However, there is no reason to expect that air pollution and noise would equally increase the risk for low birth weight, given that biological pathways would likely be different.

We simply present the findings of our study (that we found little evidence that road traffic noise is associated with birth weight after accounting for road traffic air pollution co-exposures), in the

context of the available scientific evidence, for which the volume and strength of evidence is much greater for air pollution than noise. Hypothesised biological mechanisms for both air pollution and noise are presented. Whilst biological mechanisms are not established for either air pollution or noise, the convincing evidence that maternal passive smoking during pregnancy is causally related to reduced birth weight is strongly supportive of the biological plausibility of an association between ambient air pollution and reduced birth weight, by analogy.

As to, "**why might air pollution only show an effect for birthweight, when for other outcomes such as cardiovascular health both exposures seem equally important and independent?**" We have covered this point above, where the reviewer mentions the studies by Tétreault et al., 2013; and Stansfeld, 2015. Manuscript text amended to cover this at lines 526-543.

Re the **importance of the birth weight difference** – we have covered this earlier, but in order to convey clinical significance we have altered the paper to focus on the clinical outcome term low birth weight.

The section where the results are compared to other findings is needed but it is quite long and not very easy to read for a lay reader. I think this section would benefit from some revision and cutting.
Authors' Response: We have reworked a little of this, but we do think it is necessary to put the findings into context in sufficient detail.

Limitations; the stronger findings for air pollution may be explained by differences between modelling quality for air pollution compared with noise. This should be discussed and listed as a potential limitation.
Authors' Response: In terms of model quality, both the air pollution and noise models performed similarly well in model evaluation when comparing measured vs modelled values. Air pollution: very low normalised mean bias and high spearman correlation coefficients (r) between observed vs modelled monthly concentrations: r>0.91 for NOX, PM10 and PM2.5, r>0.83 for NO2 and r>0.9 for O3 at both roadside and background locations. Noise model: high Spearman's correlation (0.90) between measured and modelled noise levels. Information on the model evaluation has been added to the Methods at lines 121 and 147. Limitations with respect to potential exposure misclassification related to the modelling are discussed at 396-416. Spatial resolution of the models differ i.e. 20m x 20m (air pollution) vs. 0.1m (noise) – this is covered above.

References

1.   Barone-Adesi F, Dent JE, Dajnak D, Beevers S, Anderson HR, Kelly FJ, et al. Long-Term Exposure to Primary Traffic Pollutants and Lung Function in Children: Cross-Sectional Study and Meta-Analysis. PLoS One. 2015;10(11):e0142565.
2.   Graham JC, C.; Barnard, M.; Mowlam, A.; McKay, S. Sole and joint birth registration: exploring the circumstances, choices and motivations of unmarried parents. Department of Work and Pensions Research Report No 463. Leeds: Her Majesty's Stationery Office by Corporate Document Services; 2007.
3.   Douglas P, Bakolis I, Fecht D, Pearson C, Leal Sanchez M, Kinnersley R, et al. Respiratory hospital admission risk near large composting facilities. Int J Hyg Environ Health. 2016;219(4-5):372-9.
4.   Pedersen M, Giorgis-Allemand L, Bernard C, Aguilera I, Andersen AM, Ballester F, et al. Ambient air pollution and low birthweight: a European cohort study (ESCAPE). Lancet Respir Med. 2013;1(9):695-704.
5.   Olsson D, Mogren I, Eneroth K, Forsberg B. Traffic pollution at the home address and pregnancy outcomes in Stockholm, Sweden. BMJ Open. 2015;5(8):e007034.
6.   Hannam K, McNamee R, Baker P, Sibley C, Agius R. Air pollution exposure and adverse pregnancy outcomes in a large UK birth cohort: use of a novel spatio-temporal modelling technique. Scand J Work Environ Health. 2014;40(5):518-30.
7.   Stieb DM, Chen L, Beckerman BS, Jerrett M, Crouse DL, Omariba DW, et al. Associations of Pregnancy Outcomes and PM2.5 in a National Canadian Study. Environmental health perspectives. 2016;124(2):243-9.
8.   Moser K, Hilder L. Assessing quality of NHS Numbers for Babies data and providing gestational age statistics. Health Stat Q. 2008(37):15-23.
9.   Basu R, Malig B, Ostro B. High ambient temperature and the risk of preterm delivery. Am J Epidemiol. 2010;172(10):1108-17.
10.   Arroyo V, Diaz J, Ortiz C, Carmona R, Saez M, Linares C. Short term effect of air pollution, noise and heat waves on preterm births in Madrid (Spain). Environ Res. 2016;145:162-8.
11.   Auger N, Naimi AI, Smargiassi A, Lo E, Kosatsky T. Extreme heat and risk of early delivery among preterm and term pregnancies. Epidemiology. 2014;25(3):344-50.
12.   Strand LB, Barnett AG, Tong S. The influence of season and ambient temperature on birth outcomes: a review of the epidemiological literature. Environ Res. 2011;111(3):451-62.
13.   Ha S, Zhu Y, Liu D, Sherman S, Mendola P. Ambient temperature and air quality in relation to small for gestational age and term low birthweight. Environ Res. 2017;155:394-400.
14.   Halonen JI, Hansell AL, Gulliver J, Morley D, Blangiardo M, Fecht D, et al. Road traffic noise is associated with increased cardiovascular morbidity and mortality and all-cause mortality in London. Eur Heart J. 2015;36(39):2653-61.
15.   Halonen JI, Dehbi HM, Hansell AL, Gulliver J, Fecht D, Blangiardo M, et al. Associations of night-time road traffic noise with carotid intima-media thickness and blood pressure: The Whitehall II and SABRE study cohorts. Environ Int. 2017;98:54-61.
16.   World Health Organisation. Night Noise Guidelines for Europe. Copenhagen, Denmark: WHO Regional Office for Europe; 2009.
17.   Stansfeld SA. Noise Effects on Health in the Context of Air Pollution Exposure. Int J Environ Res Public Health. 2015;12(10):12735-60.
18.   Tetreault LF, Perron S, Smargiassi A. Cardiovascular health, traffic-related air pollution and noise: are associations mutually confounded? A systematic review. Int J Public Health. 2013;58(5):649-66.
19.   Gulliver J, Morley D, Vienneau D, Fabbri F, Bell M, Goodman P, et al. Development of an open-source road traffic noise model for exposure assessment. Environ Modell Softw. 2015;74:183-93.

20.   Bell ML, Belanger K, Ebisu K, Gent JF, Lee HJ, Koutrakis P, et al. Prenatal exposure to fine particulate matter and birth weight: variations by particulate constituents and sources. Epidemiology. 2010;21(6):884-91.

21.   Darrow LA, Klein M, Strickland MJ, Mulholland JA, Tolbert PE. Ambient air pollution and birth weight in full-term infants in Atlanta, 1994-2004. Environmental health perspectives. 2011;119(5):731-7.

22.   Salihu HM, Ghaji N, Mbah AK, Alio AP, August EM, Boubakari I. Particulate pollutants and racial/ethnic disparity in feto-infant morbidity outcomes. Matern Child Health J. 2012;16(8):1679-87.

23.   Bell ML, Ebisu K, Belanger K. Ambient air pollution and low birth weight in Connecticut and Massachusetts. Environmental health perspectives. 2007;115(7):1118-24.

24.   Schembari A, de Hoogh K, Pedersen M, Dadvand P, Martinez D, Hoek G, et al. Ambient Air Pollution and Newborn Size and Adiposity at Birth: Differences by Maternal Ethnicity (the Born in Bradford Study Cohort). Environmental health perspectives. 2015;123(11):1208-15.

25.   Basu R, Harris M, Sie L, Malig B, Broadwin R, Green R. Effects of fine particulate matter and its constituents on low birth weight among full-term infants in California. Environ Res. 2014;128:42-51.