**Comments from the editors**

We thank the Editors for their thoughtful and constructive suggestions. We have revised the manuscript in response to the Reviewers’ comments, as detailed below.

All page/line/reference numbers refer to the clean revised manuscript.

**Please take particular note of reviewer comments below on justifying and describing the use of the Bayesian hierarchical model, on determination and clarification of confounder control, and the lack of control for potentially important factors such as BMI, smoking, and indoor air quality.**

We have addressed this, as detailed throughout relevant responses to Reviewers below. In particular please see responses to Major Comments 4, 6, 10, 15 and Minor Comments 2, 5, 8 for Reviewer 1, and responses to Major Comments 1, 2, 4, 6, 7, 8, 9, 10, 12 and Minor Comments 14, 18, 19, 21, 22 for Reviewer 2.

**Second, please ensure that the summary of main findings are fully supported by the results, e.g., take into account and explain sensitivity analyses in eFigure 1 that show weaker single-pollutant estimates for EC than those from the main models.**

We have clarified how the main findings are supported by the results throughout the revised manuscript, .e.g., (P. XX, Lines XX-XX):

*[[[QUOTE]]]*

Our main model demonstrated that there was a potentially positive association between EC and ALS diagnosis in our main model, which contained traffic-related pollutants (EC, NOx, CO) and non-EC PM2.5, as well as adjusting implicitly by matching cases and controls on age, sex, year of birth, vital status, and explicitly by including terms in the model for SES, civil status, residence, place of birth.

Because the main models contain more terms which are highly correlated, it was expected that the uncertainty of the associations would be larger than single-pollutant models. However, because of the consistency of the sign of the EC association throughout (i.e., whether in single-pollutant or multi-pollutant models), we have suggested that EC may be a driver of the relationship in the revised manuscript (P. XX, Lines XX-XX):

*[[[QUOTE]]]*

Further, we were guided by the results which showed that the single pollutant EC model (eFigure 1; model F) maintained a positive association for EC and resulted in a null association for non-EC PM2.5, while for other single-pollutant models there was still an association for non-EC PM2.5. We have added a further description in the revised manuscript (P. XX, Lines XX-XX):

*[[[QUOTE]]]*

**Finally, please ensure consistency in results presentation throughout the abstract and main text; e.g., the abstract presents posterior probabilities that are not included in the results section.**

We have ensured that there is consistency in results presentation throughout the abstract and main text, e.g., (P. XX, Lines XX-XX):

*[[[QUOTE]]]*

**Comments from reviewers**

**Reviewer #1: In this manuscript, the authors investigate the association between long-term traffic related air pollution and amyotrophic lateral sclerosis (ALS) in Denmark. The authors used Bayesian hierarchical modeling to estimate joint and individual effects of traffic related air pollutants (NOx, CO, EC) on odds of developing ALS. The major finding of this paper is that 5-year average EC concentration was individually associated with ALS. The study found no overall or joint association of traffic related pollution with ALS. This is an important topic and using methods to estimate joint effects of correlated pollutants is a necessary next step in air pollution studies. This paper was well done, though authors should explain certain analysis choices more clearly and provide more detailed results and discussion.**

We thank the Reviewer for the thoughtful and constructive suggestions. We have responded point-by-point to the Reviewer’s questions and comments below.

**Major Comments**

**1) Methods, page 5, line 46: Why did authors only include patients that were at least 20 years old at diagnosis? This choice needs to be explained/motivated.**

Cases in younger patients (i.e., less than 20 years old) would have been at a much greater chance of misclassification, with a very high likelihood that a case identified in such a young person is an error in diagnosis coding (Trabjerg et al. 2020). Further, juvenile ALS cases have been explained to much larger degree by genetic mutations (Mathis et al. 2019). We have added to the explanation of the 20-year-old limit in the revised manuscript (P. XX, Lines XX-XX):

*We only included patients who were at least 20 years old when diagnosed because (i) cases younger than 20 years old were at a greater chance of misclassification, since ALS has been predominantly diagnosed in older adults in Denmark,46 and (ii) juvenile ALS cases have been explained to a much larger degree by genetic mutations (~40%).47*

**2) Methods, page 6, line 51: Please include more detail about the spatio-temporal air pollution modeling system.**

We have now added more detail about the spatio-temporal air pollution modelling system in revised manuscript (P. XX, Lines XX-XX):

*We obtained predictions on monthly concentrations of nitrogen oxides (NOx), carbon monoxide (CO), elemental carbon (EC), and fine particles (PM2.5) (as well as ozone (O3) for a sensitivity analysis, usually negatively correlated with other pollutants due to its chemistry51), at residential addresses of study participants from the validated spatio-temporal air pollution modelling system (DEHM-UBM-AirGIS) with full space and time coverage over our study period, described in detail elsewhere.52–55 In brief, DEHM-UBM-AirGIS is a human exposure modelling system for traffic pollution, developed for application in Danish air pollution epidemiological studies. The modelling system is able to generate street configuration and traffic data based on digital maps and national databases, which enables estimation of air quality levels at a large number of addresses in an automatic and effective way.*

**3) Methods, page 7, line 17: Please provide more details on how the 1-, 5-, and 10-year averages were created for air pollution exposures. Was a weighted average created based on how long the participant lived in one location?**

We have given more details on how the 1-, 5-, and 10-year averages were created for the air pollution exposures in the revised manuscript (P. XX, Lines XX-XX):

*Based on the residential history of each case or control, we calculated 1-, 5-, and 10-year average exposure to each pollutant ending at one year before the index date, as diagnosis has been shown previously to occur at a median of 12 months after symptoms onset.60 Specifically, each case or control average value (1-, 5- or 10-year) was calculated as the mean of all concentrations recorded across time at the recorded addresses within each time window.*

**4) Methods, page 7, line 41: How did the authors determine confounding variables?**

We matched cases and controls via age, sex, year of birth, and vital status, due to ALS prevalence varying according to these characteristics. We also accounted for socioeconomic status (SES), civil status, last reported place of residence, and place of birth. SES influences many lifestyle factors, such as obesity, and has been shown as having an association with ALS diagnosis in Denmark (Dickerson et al. 2018). Civil status was included due to the influence that a spouse has on visiting a family physician (Bucher et al. 2019). Last reported place of residence was included to account for various local environmental and behavioral stressors, such heavy metals, which may have an influence on ALS prevalence (Oskarsson, Horton, and Mitsumoto 2015). Place of birth was included to account for the variety of childhood exposures, which vary by location, which may have an impact on the probability of developing ALS (Norman et al. 2013). We were also ultimately limited by what was available in the Danish Civil Registration System.

We have added these details to the revised manuscript (P. XX, Lines XX-XX):

*We included a set of covariates at index date to account for potential confounding bias, including household socioeconomic status (SES), civil status, last reported place of residence, and place of birth. We used a five-category individual-level SES definition developed by the Danish Institute of Social Sciences, based on job titles from income tax forms, which has been shown as having an association with ALS diagnosis in Denmark,61 as well as how quickly one is identified as having ALS in the Danish Civil Registration System.62 Group 1 (highest status) includes corporate managers and academics; group 2: proprietors, managers of small businesses and teachers; group 3: technicians and nurses; group 4: skilled workers; and group 5: unspecialised workers such as entry-level positions within food and retail environments. We included a group for participants who were unknown (group 9). For each married participant, we used the higher of the couple’s individual SES categories, where available. We also used information on civil status (never married, married, divorced, widowed) due to the influence that a spouse has on visiting a family physician,63 last reported place of residence from postcode (Greater Copenhagen, big cities of Denmark, rest of Denmark, Greenland) to account for various local environmental and behavioral stressors,7 and place of birth (Greater Copenhagen, big cities of Denmark, rest of Denmark, Greenland, foreign, unknown) to adjust for other potential family-specific, location-specific, and early-life confounders, which may have an impact on the probability of developing ALS.64 Ultimately, we were limited by what was available in the Danish Civil Registration System.*

**5) Methods, page 8, line 50: Please include more motivation and reasoning for ozone sensitivity analysis.**

There is evidence from other studies that ozone concentrations are associated with many different adverse health outcomes (Nuvolone, Petri, and Voller 2018). However, ozone is a highly-correlated exposure with many other pollutants included in our analysis, as can be seen from Figure 1 in the main manuscript. Adding an extra pollutant which is highly-correlated with the others may have posed challenges to the statistical inference of the model, which is why we included it in the sensitivity analysis rather than the main analysis.

Nevertheless, our sensitivity analyses demonstrated that inclusion of ozone did not noticeably change our results (eFigure 1). Different conclusions would not have been arrived at had we included ozone in our main analysis.

We have clarified why we included ozone in the sensitivity analysis in the revised manuscript (P. XX, Lines XX-XX):

*[In a sensitivity analysis, we included to account for exposures in the model, as ozone concentrations have been associated with many adverse health outcomes,69 […]*

**6) Methods, page 9-10: When discussing priors used for the Bayesian model why are weakly-informative priors given to non-EC PM2.5, but non-informative priors are given to other parameters? Please give more detail when justifying use of priors.**

We have edited the description of priors used in the Bayesian model to clarify that weakly-informative priors were given to all of the pollutant coefficients, with the other pollutant coefficients (i.e., not that of non-EC PM2.5) was given the same prior via the hierarchical structure of the traffic-related pollutant concentrations. The revised description can be found in the revised manuscript (P. XX, Lines XX-XX):

*We used weakly-informative priors so that data drove parameter estimation. Hyper-priors for coefficients on non-EC and covariates were N(0,10); for and we used Half-Cauchy(0,10), as recommended by Gelman, Polson and Scott as a weakly-informative prior;71,72 was defined by the weakly-informative prior LKJCorr(1).73*

**7) Results, page 11: Why are only 5-year average exposures presented in the results section? Please provide more justification or present the 1- and 10-year data as well in the main tables/figures.**

We provide the 5-year average exposures as the main results here because a 5-year average exposure is a balance between most recent exposure as well as long-term concentration. We have clarified this in the revised manuscript (P. XX, Lines XX-XX):

*For the main results, we present 5-year average exposure associations as a balance between most recent exposure as well as long-term concentration.*

**8) Results, page 12: Please provide more discussion of the protective effect of NOX and CO. How does this effect the null joint effect of NOX, CO, and EC?**

While the point estimates of the association of ALS and standard deviation increase for NOx and CO are less than 0% change, the credible intervals overlap with the null. We discuss further why this might be in the Discussion of the revised manuscript (P. XX, Lines XX-XX):

*[[[QUOTE]]]*

**9) Results, page 12, line 46: Please provide correlation coefficient for ozone and other pollutants.**

We now provide the correlation co-efficient for ozone and other pollutants here in the revised manuscript (P. XX, Lines XX-XX):

*O3 was negatively correlated with other pollutants, ranging from -0.58 to -0.88.*

**10) Results, page 13, line 9: Can the authors also provide posterior probability for the null?**

The posterior probability is the amount of the marginal of the co-efficient of interest which is above the null. Therefore a 50% probability means it is as likely as not that the marginal is null. We clarify this in the text with a further exposition of what various posterior probabilities mean (P. XX, Lines XX-XX):

*To calculate the probability that an association estimate was greater than null, we used the 4,000 samples of the posterior and took the proportion of samples which were above a null association. A 50% probability means that it is as likely as not that the marginal is null, a probability closer to 100% indicates that the association is more likely to be truly positive, with closer to 0% indicating more likely to be truly negative.*

**11) Results page 13, line 22: "(eFigure 1) resulted in positive associations for each of EC, NOx, CO, with positive associations for non-EC PM2.5 in all but the model with EC." Is eFigure 1 the correct figure? eFigure 1 shows protective effects for all but EC.**

We have clarified that we are only referring to the single pollutant models D, E and F in eFigure 1 in the revised manuscript (P. XX, Lines XX-XX):

*Single-pollutant models for each traffic-related pollutant adjusting for non-EC PM2.5 (eFigure 1; single traffic-related pollutant models D, E and F) resulted in positive associations for each of EC, NOx, CO, with positive associations for non-EC PM2.5 in all but the model with EC.*

**12) Discussion, page 13, line 38: Authors state that they found an average increase in concentration of traffic-related pollutants was associated with and increase in odds of ALS. Though only EC showed a positive association and joint effect was null?**

We have clarified that we are referring to the joint association here (P. XX, Lines XX-XX):

*In the largest case-control study of ALS and traffic-related air pollution to date, we found that a joint increase in average concentrations of traffic-related pollutants was potentially associated with an increase in odds of ALS diagnosis, with the clearest results for EC.*

**13) Discussion, page 14, line 48: If EC and NOX are so highly correlated why are their associations with ALS so different?**

XX

**14) Discussion, page 14, line 58: the 1-year estimate may be the most robust to exposure misclassification, provide more justification for why it may be the most relevant exposure window.**

We have added more detail in the revised manuscript (P. XX, Lines XX-XX):

*[[[QUOTE]]]*

**15) Please include justification for use of the Bayesian hierarchical model, as opposed to other mixture modeling methods (Bayesian kernel machine regression, etc.) that are more established.**

Other mixture methods, including Bayesian kernel machine regression, are not currently available for case-control studies. We have clarified this in the future research paragraph in the revised manuscript (P. XX, Lines XX-XX):

*[[[QUOTE]]]*

**16) Figure 1. Please provide figure for 1-, and 10-year exposure estimates.**

We now provide eFigures XX and XX to display the correlations of the 1- and 10-year exposure estimates.

**17) Table 1. SES group 9, why were unemployed and unclassified grouped together?**

This should have been stated as ‘unknown’ here and we have corrected this in the revised manuscript (P. XX, Lines XX-XX):

*We included a group for participants who were unknown (group 9).*

**Minor Comments**

**1) Throughout the paper there is inconsistent use of the abbreviation SD for standard deviation.**

We have now introduced the abbreviation for the term standard deviation (SD), which is also in the list of Abbreviations, we now consistently use this throughout the revised manuscript e.g., (P. XX, Lines XX-XX):

*For 5-year average pollutant concentrations, we observed the largest overall association for the individual SD increase in EC (11.5%; 95% CrI: -1.0%, 25.6%; 96.3% posterior probability of positive association per 0.42 µg/m3) (Figure 2).*

**2) Introduction, page 4, line 26: Unclear why this sentence is a contradiction.**

We are unclear about what the Reviewer means here. We would be very happy to provide a response once we understand what this comment is referring to exactly.

**3) Methods, page 8 line 42: Authors say non-EC PM2.5 adjust for other air pollutants from other sources. This would only adjust for PM2.5 from other sources.**

This Reviewer is correct and we have clarified this in the revised manuscript (P. XX, Lines XX-XX):

*If other sources of air pollution are associated with ALS, then including non-EC PM2.5 adjusts for PM2.5 from other sources.67*

**4) Methods, page 9, line 16: Model quantifies log-odds of one standard deviation increase. Please add explanation of why you chose one standard deviation increase instead of interquartile range.**

Normalizing across pollutant concentrations is necessary because we combined concentration associations via the traffic terms in the model. Both standard deviation (SD) and interquartile range (IQR) are measures of the spread of values, which can be equivalently used to normalize pollutant concentrations. There is no inherent benefit to picking one or the other in this case, as the role of dividing by both measures of spread is to normalize concentrations. We have added that both are equivalent ways of normalizing in the revised manuscript (P. XX, Lines XX-XX):

*, , , the pollutant-specific coefficients (log-odds) per standard deviation (SD) increase in concentration of , , , respectively, scaled by their respective SDs and centered at their means; and the rest as coefficients for subject-specific covariates. Interquartile Range (IQR) could equivalently be used to scale pollutant concentrations.*

**5) Results, page 12, line 4: Joint association of which pollutants?**

We have clarified which pollutants in the revised manuscript (P. XX, Lines XX-XX):

*The joint association of traffic-related pollutants (EC, NOx, CO) was 2.3% (95% CrI: -3.3%, 7.7%), with an 77.8% posterior probability of a positive association.*

**6) Results, page 13, line 17: "10-year average exposure results were attenuated versions of the 1- and 5-year results." Wording could be improved.**

We have expanded upon this phrase to be more descriptive in the revised manuscript (P. XX, Lines XX-XX):

*10-year average exposure results were attenuated versions of the 1- and 5-year results, as associations tended further to the null.*

**7) Results overall: eFigure 2 is not mentioned.**

We now mention eFigure 2 in the revised manuscript (P. XX, Lines XX-XX):

*Results from variations of the main model in the sensitivity analyses were robust to prior choices and inclusion of parish-level SES (eFigure 1). A map of average concentration of included pollutants (NOx, EC, PM2.5, CO, O3) across Denmark for a representative year (2000; middle of study period 1989-2013) is also available in eFigure 2.*

**8) Discussion, page 14, lines 29-39: If BMI is not a confounder this is unnecessary to include.**

We have deleted these lines in the revised manuscript.

**9) Figure 2. Please make the point estimate dots bigger.**

We have made these points bigger in Figure 2 in the revised manuscript, copied below for convenience:



**Reviewer #2: Comments pasted below. The uploaded review contains a figure.**

**This paper describes a study of the effect of traffic-related air pollution exposures on amyotrophic lateral sclerosis (ALS). The study, set in a very large healthcare administrative database in Denmark, has several strengths, including exposure assessments that span decades and an attempt to account for a latent period between disease onset and clinical diagnosis. The report lacks clarity and detail on important aspects of the investigation, encompassing the target estimands, validity of the ALS measurement, and sources of confounding. My concerns follow:**

We thank the Reviewer for the thoughtful and constructive suggestions. We have responded point-by-point to the Reviewer’s questions and comments below.

**MAJOR  
  
1. The interpretive distinction between the 3 estimands pursued in this study was unclear. E.g., from the abstract, "… the overall and joint association for the three traffic-related pollutants (NOx, CO, and EC), as well as pollutant-specific associations." What does "overall" mean exactly, and how does "overall" differ from "joint"? Does "joint" include interactions? Does "pollutant-specific" reflect adjustment for other pollutants? For example, how would you express the parameter estimate from each of these in words that are true to the underlying mathematics?**

We distinguished between the “joint” association of the three pollutants (i.e., percentage change in odds of ALS diagnosis with increase in each of EC, NOx, CO), and “overall” association of the three pollutants (i.e., average percentage change in odds of ALS diagnosis from each of EC, NOx, CO). We have states this in the revised manuscript (P. XX, Lines XX-XX):

*[[[QUOTE]]]*

The Reviewer is correct that the “pollutant-specific” associations reflect adjustment not just for other pollutants, but also for other covariates to account for potential confounding bias, described in more detail in the revised manuscript (P. XX, Lines XX-XX):

*[[[QUOTE]]]*

Our model also accounts for the variance-covariance structure between highly-correlated exposures and their coefficients (P. XX, Lines XX-XX):

*[[[QUOTE]]]*

**2. Related to #1, the joint association is described in these terms: "This sum quantifies the association (log-odds) with ALS of a one-SD increase in the three pollutants simultaneously." Although it is mathematically possible to compute this, how well does an increment of SD in all 3 pollutants match up with the joint distribution of these pollutants in the population? Do the concentrations vary at about the same pace? i.e., can you identify locations (or location-periods) that are 1-SD apart (or 0.5 or 0.1 SD apart, etc.) on all 3 pollutants?**

As we have described in the manuscript, the pollutants are highly correlated (P. XX, Lines XX-XX):

*[[[QUOTE]]]*

This would imply that the increase in one pollutant by 1-SD would results in a similar relative increase in the other related pollutants.

**3. Introduction: the literature review seemed cursory (e.g., "… epidemiologic and toxicological studies support several plausible biological mechanisms in association with the nervous system and neurodegeneration.15-34"). I recommend citing systematic reviews (e.g., Integrated Science Assessments from the US EPA) or using "e.g." before some of the citations.**

As per the Reviewer’s suggestion we have added “e.g.” before some of the example citations in the Introduction here (P. XX, Lines XX-XX):

*[[[QUOTE]]]*

**4. Introduction/Methods: the use of ozone in this investigation was confusing. The Introduction states, "Using three air pollutants commonly used in health studies as traffic-related emissions tracers—nitrogen oxides (NOx), carbon monoxide (CO), and elemental carbon (EC)— as well as fine particles (PM2.5) and ozone (O3), we aimed to assess whether exposure to (a) each individual air pollutant is independently associated with ALS diagnosis …," which loosely implied that ozone was a traffic-related pollutant of interest. Later, in the Methods, O3 is described as being part of "a sensitivity analysis, usually negatively correlated with other pollutants due to its chemistry." This requires more explanation. Was exposure to O3 not expected to be relevant to ALS? Are the predicted O3 concentrations inaccurate?**

We included traffic-related pollutants, (EC, NOx, CO), to investigate whether each traffic-related air pollutant, individually, jointly and overall, was associated with ALS diagnosis, while also fully adjusting for other pollutants (PM2.5 from other sources and O3) and other relevant covariates. To clarify that our main focus was the traffic-related pollutants, we have edited the last paragraph of the Introduction in the revised manuscript (P. XX, Lines XX-XX):

*[[[QUOTE]]]*

While there is little existing evidence to assess whether O3 concentrations may impact ALS diagnosis, we included a sensitivity analysis to examine whether including it made a difference to the main set of results, which it did not. We explain this in the revised manuscript (P. XX, Lines XX-XX):

*[[[QUOTE]]]*

The models we have used for air pollution prediction, DEHM-UBM-AirGIS, have good predictive accuracy for all pollutants, including predicted concentrations of O3, with average monthly correlations between measured and modelled results quoted in the revised manuscript (P. XX, Lines XX-XX):

*[[[QUOTE]]]* **5. Methods: The validity of the registry for identifying ALS cases requires more detail, as it is a fundamental aspect of this investigation. Was the validation against an in-person assessment? a neurologist's in-depth review of medical records? Did the validation compare date of diagnosis with the date of symptom onset? What were the quantified indices of accuracy? As appropriate, it could be useful to mention the potential influence of outcome misclassification (or lack thereof) on the findings and mention determinants of misclassification if known.**

We had claimed the validity of obtaining ALS diagnosis data from the Danish National Patient Register based on previous work from some of our co-authors, including the senior author of this current study (Kioumourtzoglou et al. 2015). In this work, a specialist ALS neurologist made a medical record review of registry and mortality data for factors that may have been slightly related to agreement of ALS diagnosis. This previous study found that the use of hospital discharge and death certificate records are a valuable and highly reliable tool for ALS epidemiologic studies such as our current one. We have added detail to the revised manuscript (P. XX, Lines XX-XX):

*[[[QUOTE]]]*

We have added to the Limitations in the Discussion that while the Danish National Patient Register was highly reliable for ALS ascertainment, outcome misclassification may still have been possible (P. XX, Lines XX-XX):

*[[[QUOTE]]]*

While the studied quoted above did not examine date of system onset compared with date of diagnosis, another study in Ireland found that diagnosis of ALS was found to occur at a median of 12 months after symptoms onset (Galvin et al. 2017), which we have stipulated in the revised manuscript (P. XX, Lines XX-XX):

*[[[QUOTE]]]*

Nevertheless, we have added that this may also be a limitation of the work in the Discussion of the revised manuscript (P. XX, Lines XX-XX):

*[[[QUOTE]]]*

**6. Methods: "… we removed the EC concentration from the total PM2.5 mass concentration …." How did you "remove" EC? Subtraction? Using residuals from a regression model?**

Our pollutant model for PM2.5, DEHM-UBM-AirGIS (Khan et al. 2019; Jørgen Brandt et al. 2001; J Brandt et al. 2003; Frohn et al. 2021), constructed PM2.5 concentrations by adding from specific species of pollutant, one of which was EC. We were therefore able to subtract the EC concentration from the PM2.5 concentration to obtain non-EC PM2.5 concentrations. We have added a clarification to the revised manuscript (P. XX, Lines XX-XX):

*[[[QUOTE]]]* **7. Methods, adjustment for SES/occupation: was the goal to adjust for individual-level education and/or occupation (and its attendant exposures) or to adjust for overall SES in the household? E.g., "For each married participant, we used the higher of the couple's individual SES categories." This seems to be getting at household SES. Co-habitation among unmarried couples is common in Denmark. How was this addressed?**[**https://academic.oup.com/ije/article/42/2/559/737789**](https://academic.oup.com/ije/article/42/2/559/737789)**Was it important to capture information differently from previously married people? This should come down to the construct you are trying to measure. (The SES-ALS paper cited was about individual-level occupation.**

Our goal in adjusting for SES was to adjust for occupational exposure as well as household SES, as this would be potentially related to how quickly one is identified as having ALS in the registry system. Related to this, current household SES would have influenced more the speed at which ALS diagnosis would have occurred, compared with previous marriage information. Nevertheless, we have a category of ‘divorced’ people in our civil status adjustment. We have clarified in the revised manuscript (P. XX, Lines XX-XX):

*[[[QUOTE]]]*

The Reviewer is correct to point out that co-habitation is common among couples in Denmark, which would not be captured by our analysis. We have added this as a Limitation in the Discussion of the revised manuscript (P. XX, Lines XX-XX):

*[[[QUOTE]]]*

**8. Methods, timing of covariates: for covariates whose values could vary over time (e.g., occupation, civil status, parish-level SES), what was the timing of these covariates relative to the exposure and outcome periods?**

The covariates were obtained at the index date, which we have clarified in the revised manuscript (P. XX, Lines XX-XX):

*[[[QUOTE]]]*

**9. Methods: calendar time as a course of confounding? Given that air pollutant concentrations and other determinants of ALS may have changed over time, might calendar time be a source of confounding?**

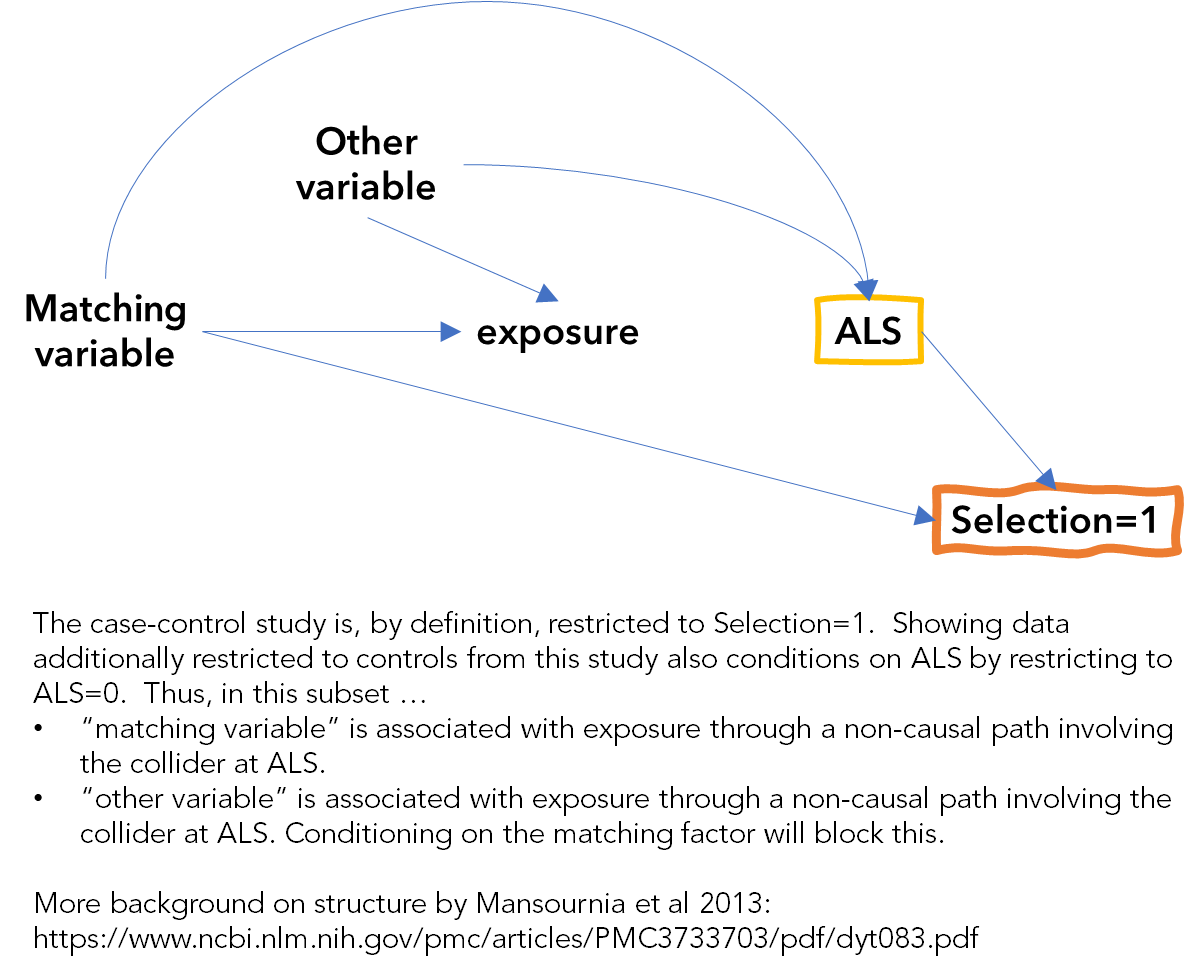
XX **10. Discussion: potential that smoking and/or BMI could be sources of confounding. The paper states, "… to induce confounding bias, any unaccounted-for variable would have to influence both ALS diagnosis and air pollution. BMI, previously associated with ALS, would not confound the association between traffic-related air pollution and ALS, as pollutant concentrations are derived independently from BMI distribution. Any BMI-air pollution association in our study, thus, would be via SES." Through complex social and economic mechanisms, the association of air pollution exposure with any given factor (e.g., BMI) can vary across populations. For example, in some study populations, areas where smoking is more common have higher concentrations of traffic-related air pollutants; in other settings, the pattern is reversed; and still others, there is little association. How is it known that BMI is not associated with exposure in this study population? Furthermore, how is it known that any such association in this study population would operate through SES? Analogous questions could be posed about smoking, as well. In the absence of firm answers to these questions, it could be useful to conduct a quantitative bias analysis, particularly as the estimated effect sizes are small.**

We agree with the Reviewer that there are complex social and economic mechanisms through which BMI may well be related to pollution levels in Demark. However, to be a potential source of confounding, BMI would have to be causing the variation in pollution levels. There is no evidence that we know of that would suggest that BMI drives variation in pollution levels. Rather, SES predicts where one lives, and also one’s BMI, and where one lives is a driver of air pollution levels, which warranted adjusting for it in our analysis.

Relevant to this, as we do not have BMI data, we acknowledge that while we do not anticipate a cause link between BMI and air pollution, and therefore no confounding, it is not testable due to our lack of BMI data for cases and controls in the Limitations section in the Discussion of the revised manuscript (P. XX, Lines XX-XX):

*[[[QUOTE]]]*

**11. Table 1/Table 2: Distribution of covariates by exposure level. To provide more information on correlates of exposure in the underlying population, conditional on the matching factors, it would be helpful to provide a table showing the co-distribution of key covariates and air pollutant exposure among the controls. This could be a challenging proposition, though. The matching scheme means that the associations of matching factors with exposure may be distorted so long as there are other common causes of exposure and ALS. Thus, it is not clear whether, without extensive additional exploration, it would be possible to show meaningful co-distributions of exposure with the matching variables. However, conditional on the matching variables, it may be informative to show associations of other covariates with exposure. See example DAG below.**



XX

**12. Methods/Bayesian hierarchical approach: I am not particularly fluent in Bayesian methods, so will leave to others to evaluate the particulars of this approach. Nonetheless, if possible but without offering an entire course in Bayesian methods, it would be helpful for readers like me to see a clearer justification for choosing this method over, say, conventional conditional logistic regression (some of this clarification might tie into clarifying the target estimands mentioned in #1), and motivation or intuition for some of the specific steps and interpretation of posterior probability. For example, in writing, "We placed a hierarchy on the traffic-specific pollutant terms in the model," does this mean that you have assumed that one pollutant emanates from another? Also, does it make sense to compute credible intervals for the posterior probabilities?**

The main advantage of the Bayesian hierarchical structure we have utilized in our analysis is that the variance-covariance structure of the traffic-related pollutants (EC, NOx, CO) can be incorporated into the model, enabling an estimate of each individual pollutant’s association with ALS diagnosis, as well as a joint (i.e., percentage change in odds of ALS diagnosis with increase in each of EC, NOx, CO), and overall (i.e., average percentage change in odds of ALS diagnosis from each of EC, NOx, CO) associations. The Reviewer is correct that this enables the model to account for the fact that the traffic-related pollutants originate from common sources. We have added this description in the revised manuscript (P. XX, Lines XX-XX):

*[[[QUOTE]]]*

The posterior probabilities are generated from the full posterior probability distributions of the marginals for each pollutant-specific, overall and joint association. To calculate the posterior probability that an association estimate was greater than null, we took a large amount of draws from these full distributions (4,000 in our case) and took the proportion of samples which were above a null association. To clarify how to interpret this value, for which there is no interpretable credible interval, we have added a description to the revised manuscript (P. XX, Lines XX-XX):

*[[[QUOTE]]]*

**MINOR  
  
13. Abstract: "For a standard deviation (SD) increase in 5-year average…." For more context, please provide the SD for each pollutant.**

We now provide the SD values for each pollutant in the Abstract of the revised manuscript (P. XX, Lines XX-XX):

*[[[QUOTE]]]* **14. How PM2.5 was used in this investigation was presented with some ambiguity. The introduction states, "Using three air pollutants commonly used in health studies as traffic-related emissions tracers—nitrogen oxides (NOx), carbon monoxide (CO), and elemental carbon (EC)— as well as fine particles (PM2.5) and ozone (O3), we aimed to assess whether exposure to (a) each individual air pollutant is independently associated with ALS diagnosis," suggesting that PM2.5 was being considered as a traffic-related air pollutant (similar to the aforementioned situation for ozone). PM2.5 is not necessarily traffic-related, as the authors later state, but here it appears to be one of the primary exposures of interest.**

To avoid ambiguity, which the Reviewer has correctly pointed out, we have removed reference to PM2.5 at this point in the Introduction in the revised manuscript (P. XX, Lines XX-XX):

*[[[QUOTE]]]* **15. A curiosity: the relative difference in odds (percentage difference in odds) is effectively an arithmetic variation on the odds ratio. Was there a particular reason the authors opted for the percentage difference expression?**

The Reviewer is correct, and we opted for percentage difference purely for wider interpretability, i.e., so a non-specialist epidemiologist might be able to read through the results and understand the numerical output.

**16. Abstract: Given the results, it was a surprise to see this conclusion in the abstract "Our results indicate a potential positive association between ALS diagnosis and pollutants, particularly for EC." Perhaps this ties into clarifying the contribution of the Bayesian approach?**

We attempted to make our Conclusions in the original Abstract indicative but not conclusive, which is the reason behind using the phrase ‘potential association’. We have further clarified with the clear statement that our results are inconclusive in the revised manuscript (P. XX, Lines XX-XX):

*[[[QUOTE]]]*

**17. Methods/Index date: Please state earlier that the date of diagnosis as indicated in the database is the index date. I.e., "We identified ALS cases based on their International Classification of Diseases (ICD) discharge diagnoses, …, using the date of the first relevant code as the diagnosis date. This was the index date."**

We have done this (P. XX, Lines XX-XX):

*[[[QUOTE]]]*

**18. Methods/matching scheme: what was the degree of match sought for age and year of birth (within months, years?).**

The matching was made by age and single year of birth. Matching by a finer scale was not possible with this dataset. We have clarified in the revised manuscript (P. XX, Lines XX-XX):

*[[[QUOTE]]]*

**19. Methods/study design: The control-sampling scheme seems to follow a risk-set matching pattern, so cases could serve as controls. If that is correct, could state that. It also means that computed ORs are estimates of IRs.**

The Reviewer is correct. We have stated this in the revised manuscript (P. XX, Lines XX-XX):

*[[[QUOTE]]]*

**20. Methods/occupational classes: these are likely official terms of the DK government, but they are not very descriptive and "unskilled" is somewhat derogatory. Although extensive detail is not needed, a little more would be informative.**

We accept the Reviewer’s point that the ‘unskilled’ by itself as a term is derogatory, so have provided some context in the revised manuscript (P. XX, Lines XX-XX):

*[[[QUOTE]]]*

**21. Discussion: "If other sources of air pollution are associated with ALS, then including non-EC PM2.5 adjusts for other air pollutants from other sources." Is it known that air pollutants that fall outside of PM2.5 (most obviously, anything in the coarse fraction of PM10) are not related to ALS risk?**

We have corrected this sentence in the revised manuscript (P. XX, Lines XX-XX):

*[[[QUOTE]]]*

**22. This phrasing was unexpected: "The conditional approach automatically accounts for matching factors (age, sex, year of birth, vital status) …." What is meant by "automatically accounts for"?**

We have clarified the language in the revised manuscript (P. XX, Lines XX-XX):

*[[[QUOTE]]]*

**23. Discussion: Please take care to avoid relying on null hypothesis significance testing to interpret the findings. See the journal's guidance here:**[**https://edmgr.ovid.com/epid/accounts/ifauth.htm/**](https://edmgr.ovid.com/epid/accounts/ifauth.htm/)**In addition, the American Statistical Association issued a strong critique of significance testing (**[**https://www.amstat.org/asa/files/pdfs/P-ValueStatement.pdf**](https://www.amstat.org/asa/files/pdfs/P-ValueStatement.pdf)**), and additional cogent arguments along these lines have been issued elsewhere, including Nature. (**[**https://www.nature.com/articles/d41586-019-00857-9**](https://www.nature.com/articles/d41586-019-00857-9)**)**

We agree wholeheartedly with the Reviewer that testing significance as the sole mechanism to deciding whether results are non-null should be avoided, as it focuses purely on the extreme part of the distribution and is prone to dismissing potentially important conclusions from results. One of the main benefits of our Bayesian approach is that each marginal’s distribution probability mass is described fully, and therefore one is able to examine how much of the probability mass is above a null association, which in our view, and many statisticians’ views, is a more ‘natural’ way of describing the confidence that a result is non-null.

We have avoided any reference to p-values, as they do not have a place in Bayesian analysis, and have now avoided any use of words related to the significance of a result in the revised manuscript, e.g., (P. XX, Lines XX-XX):

*[[[QUOTE]]]*

**\* \* \* \* \***

**Preparing a revision**

**1. For estimates of causal effects, we strongly discourage the use of categorized P-values and language referring to statistical significance, including whether a confidence interval covers the null. We prefer instead interval estimation, which conveys the precision of the estimate with respect to sampling variability. We are more open to testing with respect to modeling decisions, such as for tests of interaction and for tests for trend.**

We have avoided p-values throughout.

**2. We do not permit acronyms unless they are generally recognized by epidemiologists (e.g. HIV is okay, but LVA is not). When in doubt, we recommend that you spell out.**

We have been careful to introduce acronyms where used.

**3. Please do not include uninformative precision (excessive decimal places). For example, percents should be rounded to nn%, n.n%, or 0.0n% and risk ratios should be rounded to nn, n.n, or 0.nn unless clarity of the presentation and the sample size justify**

**more significant digits.**

We have done this.

**4. Please be sure to include explicit information about approval of human subjects research by an independent review board. If no such review was required, include an explicit statement about why the requirement for review was waived.**

We have done this in the manuscript P. XX, Lines XX-XX):

*This study was approved by the Institutional Review Board Committee at the Columbia University and the Danish Data Protection Agency.*

**5. Do not include public health policy recommendations in Brief Reports or Original Articles that present new research findings.**

We have not included any public health policy recommendations.

**6. Data appearing in the abstract must also be cited in the main text, not just in tables or figures.**

We have done this.

**7. Resubmissions must adhere to word limits. The word limits for main text (generally the introduction, methods, results, and discussion) are 1500 words for Brief Reports (plus 150 words for its abstract), 4000 words for Original Articles (plus 250 words for its abstract), 5000 words for reviews (plus 250 words for its abstract), 2000 words for Commentaries (no abstract), 600 words for Research Letters (no abstract), and 400 words for Letters to the Editor (no abstract).**

We have done this, with an Abstract of XX words and an Original Article of XX words in the revised manuscript.

**8. We advise that total word counts for Original Articles should not exceed 7500 words and for Brief Reports should not exceed 3500 words. The total word count includes main text (introduction, methods, results, and discussion), bibliography, figure legends, tables, and figures (250 words per figure, including each figure in a panel). The title page, abstract, acknowledgments, and funding information do not count in the total word count.**

We have adhered to this, with a total word count of XX words in the revised manuscript.

**9. Figure labels: Make font size as large as possible, so as to be legible when figures are reduced for publication (typically one column [8.5cm] in width).**

We have made the Figure labels large and legible.

**10. Footnotes to tables and figures should use superscript lowercase letters to link content to the footnote, not symbols or numerals.**

The footnote in Table 1 uses a superscript lowercase letter.

**11. Do not use parenthetical phrases like “(data not shown), (results not shown), or (available from the authors upon request).” In these circumstances, the data or results should be provided in Supplementary Digital Content.**

We have avoided any use of these phrases.

**12. Additional details regarding submission requirements can be found in the Instructions for Authors, which are posted at**[**http://edmgr.ovid.com/epid/accounts/ifauth.htm**](http://edmgr.ovid.com/epid/accounts/ifauth.htm)**.  
  
Preparing for resubmission**

**13. Prepare a response document for the Editor that responds point-by-point to the reviewers' comments (presenting each comment followed by your response). Give the page number where revised text can be found and, where practical, paste revised text directly into the reply document.**

We have done this.

**14. Submit versions of the manuscript with and without your changes displayed.**

We have submitted clean and tracked versions of the revised manuscript.

**15. Supplementary Digital Content should be submitted as a single PDF file, and you should use our convention - e.g. eFigure 1, eAppendix 2 - to label and refer to online content.**

We have done this.

**16. Authors should submit copies of any closely related manuscripts (published, in press, or under review).**

We do not have similar manuscripts to submit.

**17. Please revisit information about page charges and color printing charges available in the Instructions for Authors, which are posted at**[**http://edmgr.ovid.com/epid/accounts/ifauth.htm**](http://edmgr.ovid.com/epid/accounts/ifauth.htm)**.**

We acknowledge the charges on the link provided.

**18. We request that the complete revised manuscript (with all tables and figures) be completed by 05 May 2022. If you are not able to meet this deadline, please notify the editorial office.**

We have submitted before 5th May 2022.

**Resubmitting via Editorial Manager**

**19. Log-in to Editorial Manager as an author using the credentials above.**

**20. Click on the "Submissions Needing Revision" link.**

**21. To view the previous decision letter and reviewer comments, please click the blue decision term listed under the View Decision menu.**

**22. If you would like to download the previous manuscript to make revisions, click on "Download Files" under the Action menu.**

**23. To begin the resubmission: Click "Submit Revision" under the Action menu.**

**24. Proof each screen to ensure the information is still correct (the Title, Authors, etc.), then click Next at the bottom of each page.**

**25. On the Attach Files screen, select each previous submission item that you would like to carry forward to the resubmission.**

**26. Upload the revised versions of the main text (with and without tracked changes), and order them with the highlighted version first.**

**27. Upload the point-by-point reply to review.**

**28. When you are finished uploading, please click Next.**

**29. Click "Build PDF for My Approval."**

**30. Click "Go to Submissions Waiting for Author’s Approval."**

**31. Wait for the PDF to build. When it has been built, you will see the link "View Submission" in the Action menu. Click "View Submission," and open the manuscript to proof your work.**

**32. If you find problems with the manuscript, click "Edit Submission" from the Action menu. Make the required changes, and begin again at the file uploads.  
33. Once the submission is complete and acceptable, click "Approve Submission" from the Action menu.**

**34. If you have difficulty with these procedures, you may send questions to**[**timothy.lash@epidemiology-journal.com**](mailto:timothy.lash@epidemiology-journal.com)**.**

Thank you for the resubmission instructions. We have followed them.

**References**

Brandt, J, JH Christensen, LM Frohn, and R Berkowicz. 2003. ‘Air Pollution Forecasting from Regional to Urban Street Scale—-Implementation and Validation for Two Cities in Denmark’. *Physics and Chemistry of the Earth, Parts A/B/C* 28 (8): 335–44.

Brandt, Jørgen, Jesper H Christensen, Lise M Frohn, Finn Palmgren, Ruwim Berkowicz, and Zahari Zlatev. 2001. ‘Operational Air Pollution Forecasts from European to Local Scale’. *Atmospheric Environment* 35: S91–98.

Bucher, Brian T, Jianlin Shi, Robert John Pettit, Jeffrey Ferraro, Wendy W Chapman, and Adi Gundlapalli. 2019. ‘Determination of Marital Status of Patients from Structured and Unstructured Electronic Healthcare Data’. In *AMIA Annual Symposium Proceedings*, 2019:267. American Medical Informatics Association.

Dickerson, Aisha S, Johnni Hansen, Marianthi-Anna Kioumourtzoglou, Aaron J Specht, Ole Gredal, and Marc G Weisskopf. 2018. ‘Study of Occupation and Amyotrophic Lateral Sclerosis in a Danish Cohort’. *Occupational and Environmental Medicine* 75 (9): 630–38. https://doi.org/10.1136/oemed-2018-105110.

Frohn, Lise Marie, Matthias Ketzel, Jesper Heile Christensen, Jørgen Brandt, Ulas Im, Andreas Massling, Christopher Andersen, et al. 2021. ‘Modelling Ultrafine Particle Number Concentrations at Address Resolution in Denmark from 1979-2018–Part 1: Regional and Urban Scale Modelling and Evaluation’. *Atmospheric Environment* 264: 118631.

Galvin, Miriam, Rebecca Gaffney, Bernie Corr, Iain Mays, and Orla Hardiman. 2017. ‘From First Symptoms to Diagnosis of Amyotrophic Lateral Sclerosis: Perspectives of an Irish Informal Caregiver Cohort—a Thematic Analysis’. *BMJ Open* 7 (3). https://doi.org/10.1136/bmjopen-2016-014985.

Khan, Jibran, Konstantinos Kakosimos, Ole Raaschou-Nielsen, Jørgen Brandt, Steen Solvang Jensen, Thomas Ellermann, and Matthias Ketzel. 2019. ‘Development and Performance Evaluation of New AirGIS–a GIS Based Air Pollution and Human Exposure Modelling System’. *Atmospheric Environment* 198: 102–21.

Kioumourtzoglou, Marianthi-Anna, Ryan M Seals, Liselotte Himmerslev, Ole Gredal, Johnni Hansen, and Marc G Weisskopf. 2015. ‘Comparison of Diagnoses of Amyotrophic Lateral Sclerosis by Use of Death Certificates and Hospital Discharge Data in the Danish Population’. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration* 16 (3–4): 224–29.

Mathis, Stéphane, Cyril Goizet, Antoine Soulages, Jean-Michel Vallat, and Gwendal Le Masson. 2019. ‘Genetics of Amyotrophic Lateral Sclerosis: A Review’. *Journal of the Neurological Sciences* 399: 217–26.

Norman, Rosana E, David O Carpenter, James Scott, Marie Noel Brune, and Peter D Sly. 2013. ‘Environmental Exposures: An Underrecognized Contribution to Noncommunicable Diseases’. *Reviews on Environmental Health* 28 (1): 59–65.

Nuvolone, Daniela, Davide Petri, and Fabio Voller. 2018. ‘The Effects of Ozone on Human Health’. *Environmental Science and Pollution Research* 25 (9): 8074–88.

Oskarsson, Björn, D Kevin Horton, and Hiroshi Mitsumoto. 2015. ‘Potential Environmental Factors in Amyotrophic Lateral Sclerosis’. *Neurologic Clinics* 33 (4): 877–88.

Trabjerg, Betina B, Fleur C Garton, Wouter van Rheenen, Fang Fang, Robert D Henderson, Preben Bo Mortensen, Esben Agerbo, and Naomi R Wray. 2020. ‘ALS in Danish Registries: Heritability and Links to Psychiatric and Cardiovascular Disorders’. *Neurology Genetics* 6 (2).