**Comments from the editors**

**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. 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. 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 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.

**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.**

We only included patients who were at least 20 years old at diagnosis because cases in younger patients (i.e., less than 20 years old) would have been at greater chance of misclassification. We have added to the explanation of the 20-year-old limit in the revised manuscript (P. XX, Lines XX-XX):

*[[[QUOTE]]]*

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

XX

**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):

*[[[QUOTE]]]*

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

XX

**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. 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, copied below for convenience. 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):

*[[[QUOTE]]]*

**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):

*[[[QUOTE]]]*

**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):

*[[[QUOTE]]]*

**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?**

XX

**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):

*[[[QUOTE]]]*

**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):

*[[[QUOTE]]]*

**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.**

XX

**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?**

XX

**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.**

XX

**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.**

XX

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

XX

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

XX

**Minor Comments**

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

XX

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

XX

**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.**

XX

**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.**

XX

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

XX

**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.**

XX

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

XX

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

XX

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

XX

**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?**

XX  
 **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?**

XX **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.**

XX

**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?**

XX **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.**

XX **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?**

XX **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.**

XX

**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?**

XX

**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.**

XX

**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?**

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

XX **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.**

XX **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?**

XX **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?**

XX

**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."**

XX

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

XX

**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.**

XX

**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.**

XX

**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?**

XX

**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"?**

XX

**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)**)**

XX

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

**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.**

XX

**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.**

XX

**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.**

XX

**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.**

XX

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

XX

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

XX

**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).**

XX

**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.**

XX

**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).**

XX

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

XX

**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.**

XX

**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.**

XX

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

XX

**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.**

XX

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

XX

**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)**.**

XX

**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.**

XX **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)**.**