PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Exposure to traffic related particle matter and effects on lung function and potential interactions in a cross-sectional analysis of a cohort study in West Sweden
AUTHORS	Carlsen, Hanne Krage; Nyberg, Fredrik; Torén, Kjell; Segersson, David; Olin, Anna-Carin

VERSION 1 – REVIEW

REVIEWER	Mohammad Javad Zare Sakhvidi 1-Postdoctoral researcher; INSERM; UMR-S 1168; Rennes; France 2- Associate professor; School of Public Health; Shahid Sadoughi University of Medical Sciences; Yazd; Iran
REVIEW RETURNED	31-Oct-2019

GENERAL COMMENTS	It is a cross sectional study. It performed on a participants in a cohort of asthma in Sweden. Actually, the Idea to compare the association of air pollution on the lung function decline is not new. However, this manuscript had the golden opportunity to consider the pollution source specific burden. Additionally, I read the paper, it needs more elaboration specifically in term of writing style. Therefore, I cannot propose the manuscript for publication in the journal Just some comments for improvement of the manuscript. Abstract In the abstract please add exactly how much as IQR? In the abstract; do you mean Age at the time of enrollment? Introduction In Introduction line 1:4: actually, the evidences are huge in occupational settings and I am in agree with authors that it is scarce for general population therefore it would be better to rewrite the sentence for general population. In the first paragraph of introduction, it would be better to reduce the focused on in-vitro or laboratory studies. Method Please add a map from the area of catchment. Because you included data on shipping pollution, it is nice to give the authors a prospective on the area. Please add a flowchart of your study. How many cases finally you added into your analysis History of occupational exposure to inhalational hazards such as silica, cotton dust, or allergens can significantly reduce the lung function. Would e nice to do a sensitivity analysis on the specific or at-risk occupational groups.

	What is the meaning of using postcode region as a covariate in statistical analysis section? Results Please consider homogeneity in writing. Sometimes used 0.95 sometimes .95 Discussion: Actually, we should distinguish between statistical and clinical significance in the studies. In this study the authors found 0.48% (95%CI -0.89% to -0.07%) decline in the FEV. It is actually not important in term of clinical interpretation. Additionally, the important thing is mostly falling in the pattern of restriction or obstruction. It should be discussed in detail in the discussion.

REVIEWER	Dirga Kumar Lamichhane Inha University, South Korea
REVIEW RETURNED	27-Feb-2020

GENERAL COMMENTS	General comments This paper investigated the association between source-specific particulate matter (PM) and lung function and also examines the potential effect modification by gene variants (GSTP1, GSTT1, and SP-A) on this association. The authors found that exposure to traffic-related PM was associated with lung function decline. The authors also observed gene-environment interaction for the selected SNPs in relation to lung function. The detailed assessment of exposure to source-specific PM is an important strength of this study. This study showed interesting finding of interaction for some genotype with industrial PM exposure. The writing is clear, and the study is an interesting extension of published work. The study of gene-environment interaction is an important research area, and this work can provide some insight into the mechanism underlying the association observed in several epidemiologic investigations. However, the current description of the analysis of gene-environment interaction is not focused on identifying interaction patterns. I suggest revising the result for interaction analysis and also providing some discussion of proposed biological mechanisms. In addition, parts of the methods section are unclear to me and need revising. Below are issues to consider and address. Introduction 1. A brief discussion of importance of GST/SP-A polymorphisms including relevance of air pollution would be helpful. 2. The novelty of the study needs some further justification in the introduction. Does this study in a very specific population provide further insight beyond the several similar studies that have done before? Methods 1. In the study population section, the authors might want to better describe how the study subjects (N = 5216) were selected. The inclusion criteria are not specified. Were all study subjects healthy? Did you consider any other inclusion criteria? 2. In statistical methods section (P 8, L8-15), the analysis of gene-environment interaction is not clearly written. Did the authors determine the optimal SNP coding (additive, dominant or recessive) before the interaction with the air pollutants was calculated? Maybe a few more details on the interaction methods would help to make this paragraph better understandable.

	3. Could you add some texts to explain how the post-hoc analysis ("phia" package in R) works in your analysis? Results 1. Authors mentioned that Hardy-Weinberg equilibrium analysis is performed in the statistical analysis section (P 8, L6). However, they do not comment on the results of analysis. Authors may wish to comment on this analysis. 2. Have you done multiple comparison tests? If you do, would your findings keep statistical significance? 3. It is hard to understand what Figure 2 actually depicts. I therefore suggest to move the figures to the supplemental material section, and either move one or two of the tables to the main paper, or create an additional figure that shows the results of the final models. Discussion 1. In the second paragraph (P 14, L36-40), the authors do not clearly explain why the most consistent association is only found for traffic-related PM, but not for other sources of PM considered in their study. Some comment or suggestions as why these results differ should be given. 2. A potential mechanism underlying the observed interaction between SP-A1 SNPs and air pollution for lung function should be discussed. Is there any study looking at DNA methylation on that SNPs and respiratory outcomes?

VERSION 1 – AUTHOR RESPONSE

Reviewer(s)' Comments to Author:

Reviewer: 1

Reviewer Name: Mohammad Javad Zare Sakhvidi Institution and Country:
1-Postdoctoral researcher; INSERM; UMR-S 1168; Rennes; France
2- Associate professor; School of Public Health; Shahid Sadoughi University of Medical Sciences; Yazd; Iran

It is a cross sectional study. It performed on a participants in a cohort of asthma in Sweden. Actually, the Idea to compare the association of air pollution on the lung function decline is not new. However, this manuscript had the golden opportunity to consider the pollution source specific burden. Additionally, I read the paper, it needs more elaboration specifically in term of writing style. Therefore, I cannot propose the manuscript for publication in the journal
Just some comments for improvement of the manuscript.

Reply: We welcome your comments to improve the manuscript and have responded to the specific suggestions below.

Abstract
In the abstract please add exactly how much as IQR?
Reply: Thank you for pointing this out to us, this has been added (line 39).

In the abstract; do you mean Age at the time of enrollment?
Reply: Thank you for this comment. Yes, we have added this to the abstract (line 32).

Introduction

In Introduction line 1:4: actually, the evidences are huge in occupational settings and I am in agree with authors that it is scarce for general population therefore it would be better to rewrite the sentence for general population.

Reply: Thank you for this comment. We have added a statement about this to the introduction, please see line 65 onwards:

"[4] However, there is little evidence of the relevance of particles of different sizes and from specific sources to respiratory health on a population level.[5] To date particle sources have only been addressed in few epidemiological studies of respiratory health effects with non-conclusive results…"

In the first paragraph of introduction, it would be better to reduce the focused on in-vitro or laboratory studies.

Reply: We have reviewed the text and references in the introduction, but as currently there are only few epidemiological studies of effects of specific sources on respiratory health, we feel that both evidence from panel studies and in-vitro experiments is necessary to provide a background and rationale for the current study.

Method

Please add a map from the area of catchment. Because you included data on shipping pollution, it is nice to give the authors a prospective on the area.

-Reply: Thank you for this suggestions. We now refer to a map in the paper describing the exposure in detail "Segersson et al., 2017" which shows but the area and the prospective. We refrained from adding a map, because we already feel that there are a lot of figures and tables in this manuscript. Upon the reviewers' insistence, we will include a map if this is deemed absolutely imperative.

Please add a flowchart of your study.

Reply: Thank you for this suggestion. As the manuscript already has a large number of tables and figures we have described the details of the exclusion process in results chapter (line 182 and onwards),

"After excluding individuals with missing data on demographic variables such as smoking status (25), environmental tobacco smoke (76), and who had missing, or very low quality of lung function (532), there were 6006 individuals, further 333 had a missing postcode, 315 did not have a European background, and 457 were outside the catchment area leaving 5216 for the main analysis. In the genetic analysis, up to 276 individuals had missing data. Finally, 5216 were included."

How many cases finally you added into your analysis

Reply: The number of cases is stated in table 1, and we have now added them to the text in the results chapter (line 186). Additionally, we describe the population selection (see above).

History of occupational exposure to inhalational hazards such as silica, cotton dust, or allergens can significantly reduce the lung function.

Would be nice to do a sensitivity analysis on the specific or at-risk occupational groups.

Reply: We performed sensitivity analysis stratified by smoking status, atopy, asthma status, and BMI category but none showed changes or were significant (line 147).

"Stratifying data by smoking status, atopy, asthma status, and BMI category showed no significant effect modification on the estimates for air pollution effect on lung function in either linear or logistic analysis (data not shown)."

We do not have the coded information of occupational exposure available in the current data set and it is not certain that the exposed groups is sufficient to obtain significant results. However, if the reviewers insist, we can obtain the data and perform these analyses.

What is the meaning of using postcode region as a covariate in statistical analysis section?

Reply: In previous studies of respiratory health effects of air pollution, an adjustment for urban region was necessary as socio-demographic factors were seen to bias the results with affluent individuals with different risk factor panorama and health behaviors living in the most polluted areas. We found that the variable for urban region provided the best means to remove this confounding. Thus, we adopted a similar strategy, and observed a similar phenomenon where the effect estimates in a model which was not adjusted for residential region were substantially affected after municipality differences were accounted for, a sign of confounding being addressed. We have added some text in the discussion, as well as a reference to this.

> *"During initial analysis and covariate selection, we found that residential region was an effect modifier, and included this as a covariate in the study. Other studies of lung function within a single region have adjusted for municipality to avoid confounding of the results which is likely due socio-economic distribution of the study population in some urban areas, where high-exposed areas also have a high proportion of individuals with high socio-economic status which entails other risk factor panorama and health behaviours.[38]"*

Results

Please consider homogeneity in writing. Sometimes used 0.95 sometimes .95

Reply: Thank you for this comment, the manuscript has been rewritten for style.

Discussion:

Actually, we should distinguish between statistical and clinical significance in the studies. In this study the authors found 0.48% (95%CI -0.89% to -0.07%) decline in the FEV. It is actually not important in term of clinical interpretation. Additionally, the important thing is mostly falling in the pattern of restriction or obstruction. It should be discussed in detail in the discussion.

Reply: Thank you for the suggestion, we have added some discussion about this (line 245-260).

> *"In the analyses, the observed average decreases were numerically small and without individual-level clinical significance, but in logistic regression models with binary outcomes, FEV_1 below LLN was associated with high exposure to PM_{10} and $PM_{2.5}$ traffic particles, and FVC below LLN was associated with traffic particles in all size fractions as well as total PM_{BC} (Table S4). This pattern was also found when exposure was expressed categorically for a continuous outcome (Figure 1). We observed no associations with airflow limitation, rather the negative associations with exposure means that such effects, which could possibly explained by the parallel reduction of both FEV_1 and FVC. "*

Reviewer: 2

Reviewer Name: Dirga Kumar Lamichhane

Institution and Country: Inha University, South Korea Please state any competing interests or state 'None declared': None declared.

General comments

This paper investigated the association between source-specific particulate matter (PM) and lung function and also examines the potential effect modification by gene variants (GSTP1, GSTT1, and SP-A) on this association. The authors found that exposure to traffic-related PM was associated with lung function decline. The authors also observed gene-environment interaction for the selected SNPs in relation to lung function. The detailed assessment of exposure to source-specific PM is an important strength of this study. This study showed interesting finding of interaction for some

genotype with industrial PM exposure. The writing is clear, and the study is an interesting extension of published work.

Reply: Thank you for this kind summary.

The study of gene-environment interaction is an important research area, and this work can provide some insight into the mechanism underlying the association observed in several epidemiologic investigations.

However, the current description of the analysis of gene-environment interaction is not focused on identifying interaction patterns. I suggest revising the result for interaction analysis and also providing some discussion of proposed biological mechanisms. In addition, parts of the methods section are unclear to me and need revising. Below are issues to consider and address.

Reply: Thank you for your comments.

Introduction

1. A brief discussion of importance of GST/SP-A polymorphisms including relevance of air pollution would be helpful.

Reply: This is addressed in the introduction (line 85-88):

"Glutathione S-transferase (GST) are involved in metabolizing reactive oxygen species to reduce oxidative stress.[16] GSTP1 SNPs have been reported to modify the risk of cardiovascular disease associated with exposure to NO_2[17] and to modify the association between NO_2 and lung function decline in adults,[18] but findings are inconsistent and no meta-analysis has been performed.[19,20]"

2. The novelty of the study needs some further justification in the introduction. Does this study in a very specific population provide further insight beyond the several similar studies that have done before?

Reply: Generally speaking, there are not that many population based studies of lung function, and, to our current knowledge, and none with source-specific exposure estimates.

We have added some further discussion of this and stress it in the introduction (line 74) while at the same time acknowledging the possible shortcomings of the exposure assessment (line 261 onward)

Methods

1. In the study population section, the authors might want to better describe how the study subjects (N = 5216) were selected. The inclusion criteria are not specified. Were all study subjects healthy? Did you consider any other inclusion criteria?

Reply: We thank you for this comment and opportunity to specify our inclusion criteria, see the results chapter (line 182-188).

"After excluding individuals with missing data on demographic variables such as smoking status (25), environmental tobacco smoke (76), and who had missing, or very low quality of lung function (532), there were 6006 individuals, further 333 had a missing postcode, 315 did not have a European background, and 457 were outside the catchment area leaving 5216 for the main analysis. In the genetic analysis, up to 276 individuals had missing data. Finally, 5216 were included with information on the variables related to exposure and health outcomes used in this study and self-reported European ancestry."

We have now also specified the number of individuals with asthma (self-reported current asthma and MD diagnosed asthma) and atopy (Table 1, page 10d).

2. In statistical methods section (P 8, L8-15), the analysis of gene-environment interaction is not clearly written.

Did the authors determine the optimal SNP coding (additive, dominant or recessive) before the interaction with the air pollutants was calculated? Maybe a few more details on the interaction methods would help to make this paragraph better understandable.

Reply: We analyzed all SNPS with a dominant model. There are several reasons for this. The minor allele dominant model yields good statistical power in most situations, comparing two groups of reasonable size and this model also offers straightforward interpretability with a simple 2-group interaction. In addition, for most SNPs, the minor allele is rare and the rare allele homozygotes are very rare and contribute little to the analysis and have a large uncertainty in their estimate, meaning the minor allele dominant model and the additive model are very similar in terms of power and effect estimates. To avoid multiple testing, we used the a priori defined dominant model. We acknowledge that for the four SNPs with the lowest major allele frequency (rs 596603, 2rs76803, rs11346450, rs1059046) the use of the minor allele dominant model may be questioned, and other models might usefully be explored (at multiple testing penalty), but in the case of a rare minor allele, determining if dominant or additive effect is most appropriate is not easy. However, when we checked the SNPs with more frequent minor alleles we did not find minor allele recessive effects on our outcome, and thus refrain from investigating interactive effects on these SNPs.

We have added some clarification in the methods (line 232 onwards).

3. Could you add some texts to explain how the post-hoc analysis ("phia" package in R) works in your analysis?

Reply: Thank you for this comment. We have refrained from using this analysis and report crude interaction plots as they are more readily interpretable (see Figure 2).

Results

1. Authors mentioned that Hardy-Weinberg equilibrium analysis is performed in the statistical analysis section (P 8, L6). However, they do not comment on the results of analysis. Authors may wish to comment on this analysis.

Reply: A summary of this analysis has been added to the results section (line 229). One SNP was slightly out of HWE, but we considered this likely to be due to chance as the minor homozygotes were extremely few, and included it in the analysis. Since this SNP did not produce any notable results we do not discuss this aspect further, but prefer to keep it in the analysis for completeness.

2. Have you done multiple comparison tests? If you do, would your findings keep statistical significance?

Reply: No, as we have a plausible biological model, we considered that multiple test correction was not appropriate. However, we acknowledge the issue, and we have reduced the number of tests by restricting ourselves to the minor allele dominant model only, and present our results nominal p-values only for this exploratory analysis. In the text, we have clarified the exploratory nature of this analysis (line 238 and 245).

3. It is hard to understand what Figure 2 actually depicts. I therefore suggest to move the figures to the supplemental material section, and either move one or two of the tables to the main paper, or create an additional figure that shows the results of the final models.

Reply: Thank you for your comment. We have redrawn these plots as unadjusted interaction plots of the associations with standard error bars (figure 2).

Discussion

1. In the second paragraph (P 14, L36-40), the authors do not clearly explain why the most consistent association is only found for traffic-related PM, but not for other sources of PM considered in their study. Some comment or suggestions as why these results differ should be given.

Reply: Thank you for this comment, we have now made this clearer in the abstract (line 47) and the discussion (line 261 and onwards).

2. A potential mechanism underlying the observed interaction between SP-A1 SNPs and air pollution for lung function should be discussed. Is there any study looking at DNA methylation on that SNPs and respiratory outcomes?

Reply: Thank you for this comment. We have looked at genotypic effects and have not considered methylation at all. We think that a discussion of epigenetic phenomena leads too far beyond the core focus of this paper which is on air pollution effects with an exploratory analysis looking at some potential genotypic effects and interactions as an addition.

VERSION 2 – REVIEW

REVIEWER	Mohammadjavad Zaresakhvidi Yazd Shahid Sadoughi University of Medical Sciences, Iran Inserm, France
REVIEW RETURNED	24-Apr-2020

GENERAL COMMENTS	Thank you very much for giving me the opportunity to review the manuscript again. Unfortunately, there is no response to the reviewer's file. Authors just revised the paper without a point by point description (please let me know if I am wrong). I was able just to see a PDF file of a possibly revised manuscript. For me, it is necessary to access to the point by point response to reviewer(s) comments, I can not approve the consideration/rejection of my previous comments (with the rationale behind it). Therefore, In this form, I reject the manuscript. I would be more than happy to review the manuscript if the author resubmits it with a point-by-point description of the comments.

REVIEWER	Dirga Kumar Lamichhane Inha University College of Medicine, South Korea
REVIEW RETURNED	06-May-2020

GENERAL COMMENTS	The authors have addressed my concerns, and I recommend publication in the journal.

VERSION 2 – AUTHOR RESPONSE

- We note that reviewer 1 has not read the point-by-point response provided. We were unfortunately unable to contact the reviewer for further comment, so the response has been assessed in-house. In your response to reviewer 1's comments we see that you have carried out a sensitivity analysis which showed no significant changes. Rather than stating "data not shown" in the manuscript, we request that this data is included as supplementary information.

Reply: we have amended the text from the original statement:

"Stratifying data by smoking status, atopy, asthma status, and BMI category showed no significant effect modification on the estimates for air pollution effect on lung function in either linear or logistic analysis (data not shown)."

To a new version (page 14):

"Analysing the data stratifying by smoking status, atopy, asthma status, and BMI category showed no significant effect modification on the estimated effect of $PM_{2.5}$

from traffic sources on lung function in either linear or logistic analysis. Although the estimated effect of exposure differed between the subgroups, all confidence intervals overlapped (Table S7).”

And added the analysis as supplementary table S7 in the supplementary tables file, page 9-10.

- The strengths and limitations section after the abstract should contain 3-5 bullet points, not longer than one sentence each, that relate specifically to the methods. Please shorten the final bullet point.

Thank you for pointing this out to us, we have changed the last bullet point from

“Full residential history was not available, so exposure is only assigned for the time of inclusion into the study, which also does not take indoor or occupational air pollution into account. Nevertheless, the population is known to be relatively stable and home address exposure is commonly used as the main exposure measure for air pollution since this is where people spend most of their time.”

To a new version which also corrects some inaccuracies.

“A full residential history was not available and thus exposure was assigned for the time of the participation and indoor or occupational air pollution exposure was not taken into account.”

.

- On the online submission form you have included the following Patient and Public Involvement statement "Patients and the public were not involved in the design, or conduct, or reporting of the present research". Please ensure that this statement is also included in your manuscript. It should be placed at the end of the methods section under the heading "Patient and Public Involvement".

Thank you for pointing this out to us, the statement has been added (page 8)

“Patient and public involvement

Patients and the public were not involved in the design, or conduct, or reporting of the present research.“

Furthermore, I have changed the mailing address of co-author Fredrik Nyberg to match his current address, amended a clumsy wording in the footnotes to table 1, page 10,