Author's response to reviews

Title: Impact of air pollution on pulmonary function and respiratory symptoms in children during different seasons of the year. A cohort study.

Authors:

Benigno Linares (lisb700705@yahoo.com.mx)
Juan M Guizar (jmguizar@prodigy.net.mx)
Norma Amador (norma.amador@imss.gob.mx)
Alfonso Garcia (algavela@hotmail.com)
Rogelio Perez (perezpad@servidor.unam.mx)
Rocio Chapela (rociochm@iner.gob.mx)
Victor Miranda (vmsoberanis@gmail.com)

Version: 3 Date: 8 May 2010

Author's response to reviews: see over
Scott Edmunds PhD
Senior Editor
The BioMed Central Editorial Team

We enclose the third version of the manuscript number 1571216295273858 entitled “Impact of air pollution on pulmonary function and respiratory symptoms in children. Longitudinal repeated-measures study”.

We thank the reviewers for their thoughtful comments. We have addressed each comment specifically and to facilitate the lecture we are presenting the reviewers’ comments in bold characters and our response in low characters.

Reviewer 1.

**Major Compulsory Revisions**

The description of statistical modeling procedures remains unclear. It appears that there were 3 types of dependent variables: prevalences (tables 3 and 4), lung function levels without respect to elapsed time (“non-temporal” lung function, table 5), and lung function growth over time (table 6). Clear descriptions should be presented separately for the types of models used for assessing each of these 3 types of dependent variables. These descriptions should include clear listings of dependent variables, independent fixed effects, and random effects.

Regarding the models, we used two different types:

Generalized linear models to relate pulmonary function in children (table 5). In this case showed results are unadjusted.

In table 6, we are presenting the percent changes on lung function growth in children. Based on our data we used a three-level model as well as the proc mixed in SAS. In the first level we modeled the individual characteristics of children per visit, like age. In the second level we modeled the characteristics at community level and on the third level we modeled the characteristics measured over the period of analysis, like the mean average of pollutants. We assumed age as a random effect whereas the monitor station and child as fixed effects. Moreover, we assumed an unstructured covariance structure, and due to each child was assigned an exposure of the same monitor, we assumed that Child was nested on Monitoring station.

Statistical procedures should be described in separate paragraphs for the different types of dependent variables. I understand that the journal does not have a word limit, so these descriptions can be as long as they need to be to provide sufficient clarity.

We thank the reviewer for this suggestion and have done this change.
I also wonder whether such a complex analysis was really necessary. The authors should make every effort to describe the analysis in a way that non-experts in statistics can understand.

We thank the reviewer for this suggestion and have done this change.

It appears that table 5 presents associations of dependent lung function variables with independent variables, without respect to temporal change ("non-temporal" lung function). Were these results obtained with the random-intercept models referred to in the methods?

Yes, these results were obtained with the random-intercept models. Results are showed considering single-pollutant models, and all the listed independent variables were included. The age was considered as a random effect and school as fixed effect.

What specific intercepts were modeled as random effects? It appears that table 6 presents results of temporal growth models for lung function. Were these results obtained with the random-slope models referred to in the methods. What specific slopes were modeled as random effects? Were they slopes of lung function over time, of lung function with age, or something else? Again, these and other points should be clarified in an improved description of statistical methods.

Due to the assumption that he follow up is short enough (3-months) that a straight line adequately represents the growth curve of Lung Function vs Age, our specific intercept was the age which was modeled as a random effect. Results from table 6 were obtained with the random slope models described in the methods. The slopes modeled as random effects were of lung function on age.

Moreover, we examined random intercept models, including both, age and sex as fixed effects, against random intercepts and slopes models, including age, gender, and the interaction terms age*pollutant*gender; and age*pollutant as fixed effects and age as a random effect. For the four lung function parameters (FVC, FEV1, FEF25-75% and FEV1/FVC), we obtained the best goodness-of-fit statistics, smaller Akaike’s Information Criterion (AIC), and a variance of the random slope not equal to zero. A statistically significant interaction term between age and sex was obtained as well, which indicates that the three lung function parameters (FVC, FEV1, FEF25-75% and FEV1/FVC) intercept, and rate of change over age differ by gender.

We have added some more detailed information in the methods section.

Many types of models are described in the methods section, but the models that provided results presented in tables 4, 5, and 6 are not specified. Such specification is necessary.

We agree with this comment and have added the models on the methods section.

The presentation of the basic study setting should be changed. Specifically, air pollution is said to be higher at school 1 than school 2. But the data presented suggest otherwise. When the 5 pollutants O3, SO2, CO, NO2, and PM10 are considered, pollutant levels were higher at school 1 in 10 of 20 pollutant-season-specific comparisons, and were also higher at school 2 in 10 of 20 comparisons (table 2). Thus, it is appropriate to say that school 1 was closer to major stationary air pollution sources than was school 2, but not to say that school 1 had higher levels of measured pollutants than did school 2.
We agree with this comment and eliminated all statements meaning higher pollutant levels in school 1 than in school 2.

Prevalence rates were higher at school 1 in 20 of 24 possible outcome-season-specific comparisons, and were higher at school 2 in only 4 of 24 comparisons (table 3). Furthermore, in table 4, positive associations of air pollution with prevalence were observed in 13 of 24 possible comparisons, and negative associations were observed in 11 of 24 comparisons, a virtual tie. Thus, prevalence were more strongly associated with school than with air pollution levels. I think that unmeasured characteristics associated with school may well have been more responsible for the observed results than was air pollution. That is, to my eye, there was an overall ecological association of prevalences with school, and differences in air pollution levels between schools were insufficient to explain this association.

In discussion and conclusion it was specified that students from a school closer to the major stationary air pollution sources had in general more respiratory symptoms than those from a distant school. However, pollutant levels by itself are insufficient to explain this association.

On balance, I think that the authors overinterpreted the adverse effect of air pollution in the discussion, at least with regard to symptom and illness prevalence. The discussion should be revised accordingly.

We agree with this point and have added some detailed information on the discussion.

Also, air pollution levels were apparently more strongly associated with lung function than with symptom and illness prevalences. This point should be addressed in the discussion. How might air pollution affect lung function without appreciably affecting symptom and illness prevalence?

We agree with this point and have added some detailed information on the discussion.

Minor Essential Revisions

What method was used to measure PM10?

By TEOM

In table 2 the units of CO are wrong.

Co units were ppm

The authors' cover letter states that the baseline survey occurred in winter, and that school 2 was the reference school in statistical models. However, these points are not mentioned in the revised manuscript.

This point was added

Are the results presented in table 3 adjusted or unadjusted?

Unadjusted
Table 3: Did wheeze prevalence really exceed 90% in winter? This is rather hard to believe. Were the authors really measuring wheeze, or were they measuring a broader complex of symptoms that could have included dyspnea without wheeze. ARTI and hospitalization for ARTI are not symptoms. In tables 3 and 4, ARTI and hospitalization for RTI should not be separated by dyspnea.

These data is part of the ISAAC survey and was not confirmed by the investigators.

Tables 4 and 5: Are these findings from single-pollutant models of multi-pollutant models? Were all of the listed independent variables included in single models? Why is CO not included in tables 4 and 5? Was CO excluded from analysis? If so, why?

We eliminated this variable from the analysis because the high number of missing data.

In tables 4 and 5, what is the meaning of the independent variable for fossil fuel? Is this variable really helpful? The authors state that some gaseous pollutants are associated with fossil fuel. So is the fine fraction of PM10, which usually constitutes the majority of PM10.

Use of fossil fuel (including charcoal and woodcutter) frequently used in Mexico was considered.

In the footnotes of table 6, specify the pollutants for which IQRs are given. It is stated that lung function values were log-transformed. What is the proper interpretation of the coefficients presented in table 6? Are they antilogs of the actual model coefficients?

Yes, the coefficients are antilogs of the actual model coefficients and represent the percentage change in lung function according to the IQR.

Discretionary revisions

The authors mention the National Institute of Occupational Safety and Health. Is this the NIOSH in the U.S. or some other NIOSH? Yes it is.

It appears that only measurements from stationary air pollution monitoring stations were used. If this is true, there is no need to mention the mobile monitoring stations.

We eliminate the mobile monitoring station mention in methods.

I don’t understand why season was not modeled in the revised statistical analysis. Maybe the authors misunderstood my initial comment. As I stated before, I think that the main purpose of the paper is indeed to evaluate air pollution effects, not seasonal effects. Even so, adjusting models for season would be appropriate, even though evaluation of season is not the main purpose of the paper. I think that season could be included in the models and fossil fuel could be excluded. (Admittedly, this model specification might lead to confounding of season with temperature.)
We thank the reviewer for this comment. As we stated, data was collected through a 12-months period by getting two measures per season (warm and cold), therefore, if we include the seasonal term we don’t have enough measures in order to assess the change in lung function growth, since we have children with only one measure.

We have addressed this issue in the discussion section.

**Why was temperature consistently about 4 degrees higher at school 1 than school 2 (table 2)? Can this difference be briefly explained?** We have not a specific explanation. However school 2 is near a river and has more vegetation.

**REVIEWER 2**

**Major Compulsory Revisions**

As a statistician, I have only valued more methodological points and techniques used, but I think a good article. The researchers are engaged and conclusions are of social importance.

1) In general need to improve the statistical methodology section as it is very general and unspecific.

We thank the reviewer for his thoughtful comment. We have included some more detailed information about statistical methodology in the Methods section.

2) The sample is small as in other cohort studies have been made bigger, this inconvenient can affect the power of the study and affect to the negative test p-values, this can be taken into account in the conclusions and discussion. It is necessary to justify the sample size to 15% parameter which indicates to detect differences in respiratory symptoms. In what is based on? Should indicate the test with which you have calculated the sample size (using binomial proportions test?)

In the corresponding section, this point was specified.

2) On paragraph (Statistical Analysis), is not known whether there had been some discussion about the normality of continous variables (Kolmogorov test of goodness of fit?). Should be conducted and provide data on these outcomes.

We agree with this comment. We did test the normality of response variables: FVC, FEV$_1$, and FEF$_{25-75%}$ via the Shapiro – Wilk test for goodness of fit. Data did now show a good fit, hence we used the log-transformed version of the spirometry tests values.

Some more detailed information has been added on the methods section.

Moreover the statistical methodology section is confusing and not well understood how they have performed the analysis. I believe that due to the lack of a better explanation and redaction of it. Should be use a better redaction and explain what were the models used; in the abstract indicates that it is a longitudinal study of repeated measures in time and have been analyzed using mixed models, but also is cited a paragraph very confusing, "Multilevel models (....” during the statistical analysis paragraph, please clarify.
We agree with this comment and have added information on Methods as well as on the Abstract.

Should also clarify the type of statistical analysis model used as the mixed model and what are the factors analyzed (age, sex, individual, time, etc ...) they have been taken into account repeated measures in time and post-hoc tests been made to compare levels of the factors. How many levels does each factor and scale is used ...

Based on our data, we used a three-level model with repeated measures as follows:

Let $Y_{cij}$ denotes the lung function value for subject $i$ at visit $j$ in community $c$ ($c = 1,2$). Let $t_{cij}$ denotes the age of subject $i$ at visit $j$ in community $c$. $Z_c$ denote the pollutant average in community $c$. Assuming the follow up is short enough (3-months) that a straight line adequately represents the growth curve of Lung Function vs Age. Hence, we have a three-level model as follows:

1.- $Y_{cij} = a_{cij} + b_{cij} (t_{cij}) + e_{cij}$
Where $a_{cij}$ denotes a subject-specific intercept, $b_{cij}$ is a subject-specific slope, and $e_{cij}$ are the residuals assumed to be normally distributed.

Second stage:

2a.- $a_{cij} = a_c + \beta_1 \text{Acute}_\text{poll} + e_{ci}$                                                   2b.- $b_{cij} = b_c + \beta_2 \text{Acute}_\text{poll}$
+ $f_{ci}$
Where $\text{Acute}_\text{poll}$ denotes the 3-months average exposure per visit. Residuals $e_{ci}$ and $f_{ci}$ are multivariate normally distributed, whereas $a_c$ and $b_c$ are town-specific averages of the intercepts and slopes respectively.

Third stage:

3a.- $a_c = \alpha_1 + e_c$                                                              3b.- $b_c = \alpha_2 + f_c$

To get either two-pollutant or three-pollutant model, just add the corresponding pollution terms on models 1 to 3b.

Covariates: age, BMI, and BMI squared.

Finally in this section does not mention anything about the logistic regression used later. Should indicate what were the significant factors identified, as calculated by the OR and its confidence interval (method of Wolf).

We specified statistical analysis as suggested
Sincerely Yours

Dr. Juan Manuel Guízar
Corresponding author
jmguiz@guidy.net.mx