Reviewer's report

Title: A study protocol to evaluate the relationship between outdoor air pollution and pregnancy outcomes

Version: 1 Date: 16 July 2010

Reviewer: Denis Zmirou-Navier

Reviewer's report:

General comments:

This paper describes the design of a retrospective semi-ecological study aiming at assessing the influence of exposure of pregnant women to the complex make-up of outdoor air pollution on pregnancy outcome measured as low birth weight and preterm birth. The study location is the littoral zone of the Alentejo region, south of Portugal, a mix of urban, including industrial, and rural/agriculture environments. The main focus in this paper is put on exposure assessment, which is the parameter assessed at an ecological (geographical) scale. The pregnancy outcome data being assessed at an individual scale through primary health care services and personal interviews. The originality of the study is to rest on a global indicator of air quality, the lichen diversity index value (LDV), which is the conjunction of lichen species richness and abundance, assessed in 98 different locations chosen so as to cover the whole 5300 square kilometers area and its geographical diversity. Using this indicator, that will allow to derive a continuous metric of exposure, the authors ambition to improve the assessment of exposure of each pregnant women (whose personal address during pregnancy is to be geocoded) compared to what other papers in the literature have proposed, based on a limited number of air quality monitors. Lichens have indeed been shown to be well correlated to air quality in industrial and urban settings.

My major concern deals with the fact that this correlation has been shown for long term values of air concentrations for a variety of pollutants (SO2, metals, PAHs...), not to shorter term variations. Now, the impact of air pollution on pregnancy outcome is to be mostly (maybe not entirely, however) expected from exposure during the pregnancy period itself, and specifically during certain vulnerable periods of the pregnancy (first, second or third trimester according to the outcome of interest). This time assessment of relevant exposure cannot be correctly captured using lichens which, as correctly stated by the authors, develop and accumulate symptoms of their own exposure along years. Air pollution contrasts between the study period locations might well be smaller than the seasonal (winter/summer) variations associated with emissions and dispersion conditions throughout the year and therefore the lack of appropriate time resolution of the exposure metric might induce a high degree of misclassification, when translated at the individual level.
These issues would at least warrant an extensive discussion which is completely absent. Also, the spatial resolution is of concern. The grid chosen to collect lichen samples are 4km wide squares. This is quite large when it comes to capture effects of traffic emissions or small stationary sources of pollution (e.g., small plants) where the decay of pollutants concentrations is steady within 10th to few 100th of meters, not kms. Again, short distances variations of pollution (NO, benzene…) may be larger than the contrasts that will be captured by the lichen index, smoothed across the study region by the statistical model the authors say they will use.

Major Compulsory Revisions

- Introduce an extensive discussion on the pros and cons of the exposure assessment approach, in particular in reference to the temporal and spatial resolution issues exposed above.

- p 13: Regarding confounding factors, no information seems to be collected on occupation and associated exposures (both chemical [usage of pesticides in agriculture; working at petroleum plan …] and physical risk factors: standing position etc) nor on personal smoking. Please comment.

- Also add a short section about possible selection bias linked to participation acceptance and its likely impact on risk estimates.

Minor Essential Revisions

Methods section:

- Page 10, the authors state that the choice of the Lichen sampling spots avoided “sites with local disturbances, particularly main roads or other facilities like farms or small factories”. This might result in underestimation of the local-short distance pollution variability, which is much greater than long term- and wide scale average differences. Add a comment.

- Page 13: it is said that information on “residential proximity to air pollution sources” is collected through questionnaires. Now air pollution is already assumed to be captured by the lichen index. Comment on whether this might or not produce a redundant information.

- P 13, the section on the statistical analysis (logistic regression) is unclear.

Discretionary Revisions

- Page 5: what do the authors mean about the air quality sampling sites being sit at “much higher altitude ?”. They are typically between 2 to 4 meters high. This did not disqualify them to assess population exposure in hundreds of published studies.

- The outcomes relate only to babies born; the authors might want to add a word about the fact that still births are not included, thus underestimating the possible impact of air pollution.

- Page 10: what does “the same type of land management” implicate in terms of
sampling criteria?
- P 13, What information is collected about diet as a possible confounder or effect modifier (remove the term “effect modifier” at the entry of this paragraph and substitute by the more neutral term of “other risk factors”) ?
- The last section, before a short conclusion, might describe where the study currently stands, in order to provide the reader with a sense of how the proposed protocol will indeed be undertaken

**Level of interest:** An article whose findings are important to those with closely related research interests

**Quality of written English:** Acceptable

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

I declare that I have no competing interests