Reviewer's report

Title: A Global Perspective on Coal-fired Power Plants and Burden of Lung Cancer

Version: 0 Date: 10 Sep 2018

Reviewer: Stefano Parodi

Reviewer's report:

The manuscript by Lin and collaborators provides an estimate of the excess of lung cancer incidence associated with coal combustion using a very large data set with measures taken at national level. It is an interesting and well written paper, statistical analyses are (in general) correctly applied, and the provided estimates of attributable risk are potentially useful both for health policies and for research purposes. However, in my opinion, the Authors have overemphasized the relevance of their results. In particular, estimates obtained from an ecological regression (even from a sophisticated model using GEE) should be considered cautiously due to the ecological bias. I deem that the sentence "the potential for ecological fallacy is unlikely because our analysis on aggregated data is meant to infer policy decisions at the national level" (Study Limitations paragraph, page 14) is too optimistic.

Major revisions:

The Authors should discuss the validity of their findings at the light of the potential effect of the ecological bias and uncontrolled confounding, especially from occupational exposures.

Minor revisions:

1) The excess risk of lung cancer associated with coal combustion is attributed to the release of particulate matter only, whereas a lot of different pollutants, including carcinogenic compounds, are produced during the coal combustion.

2) Abstract, page 2, rows 22-23: the Authors should put less emphasis on their results. An ecological study cannot "demonstrate" a relationship; it can just suggest that an association exists.

3) Introduction, page 4, rows 7-9: particulate matter is associated with cardiovascular risk also in subjects without cancer.
4) Data Analysis, page 7, rows 8-12: GEE models can provide biased estimates of an association if data are prone to some bias (e.g., uncontrolled confounding, or ecological fallacy).

5) Falsification Test, page 8, row 1: codes C18 to C21 are correct, but (at least to my knowledge) "1030 code" does not belong to the ICD10 classification.

6) Burden of diseases analysis, page 8, row 4: PAF indicates the population attributable "fraction".

7) Table 1: "incidence" indicates incidence rates per year? At page 9 an incidence rate of 45.68 per hundred thousand is reported that does not correspond to the results in Table 1.

8) Table 1: in my opinion confidence intervals for variables measured at a national level as the number of males are not meaningful. I also strongly suspect that 95% CI for such estimates are too large.

9) Results, page 10, row 22: relative risk of lung cancer for males corresponds to that of females in Table 2 and vice versa. Furthermore, confidence intervals and point estimates should be consistently reported, e.g., 85% (95% CI=22% - 182%), or 1.85 (95% CI=1.22 - 2.82).

10) Table 2: "adjusted for different variables in different models" should be specified in the footnotes, and not in the title.

11) Supplemental Table 3: please check for the correctness of the "RR" empty columns.

12) Discussion, page 15, row 17: "smoking is unlikely to be a confounder at national level (due to lack of association with coal capacity)". It is true, but smoking habit is often associated to socioeconomic conditions that can be strongly related to environmental exposures (polluted areas are often associated to high level of deprivation). Accordingly, smoking habits could have contributed to ecological fallacy, together with uncontrolled confounding from occupational exposures.

13) Please check the correctness of the authorship in citation 26 (Liang & Zeger).
**Level of interest**
Please indicate how interesting you found the manuscript:

An article of importance in its field

**Quality of written English**
Please indicate the quality of language in the manuscript:

Acceptable

**Declaration of competing interests**
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organisation that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I declare that I have no competing interests

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal