Author’s response to reviews

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

Authors:

Cheng-Kuan Lin (chl309@mail.harvard.edu)
Ro-Ting Lin (roting@mail.cmu.edu.tw)
Tom Chen (tomchen00@gmail.com)
Corwin Zigler (czigler@hsph.harvard.edu)
Yaguang Wei (weiyg@g.harvard.edu)
David Christiani (dchris@hsph.harvard.edu)

Version: 1 Date: 22 Oct 2018

Author’s response to reviews:

Reviewer #1: This manuscript presents a geographical ecological analysis of the contribution of coal-fired power plants to lung cancer risk, along with estimates for other contributions to lung cancer risk. The geographical ecological approach is well-suited to investigating such associations. When done at the global scale, it is similar to using satellite images to see the big picture.

Response: We would like to thank you for reviewing our manuscript and providing constructive suggestions, which help improve the quality of this manuscript. We have carefully revised the manuscript and our responses are provided as below.

Suggest mentioning that PM2.5 appears to be linked to all of the cancers for which smoking is a risk factor:

Air pollution in relation to U.S. cancer mortality rates: an ecological study; likely role of carbonaceous aerosols and polycyclic aromatic hydrocarbons.

Grant WB.

Anticancer Res. 2009 Sep;29(9):3537-45.
Response: Thank you for your suggestion. We added the references after sentence “However, air pollutants emitted from coal power plants and their potential impact on population health have aroused widespread concerns; fine particulate matter (PM2.5) can cause both short-term and long-term adverse health outcomes (2-4)”.

Here is an ecological study of particulate matter and ischemic heart disease.

Particulate air pollution and chronic ischemic heart disease in the eastern United States: a county level ecological study using satellite aerosol data.

Hu Z, Rao KR.

Spatial analysis of MODIS aerosol optical depth, PM2.5, and chronic coronary heart disease.

Hu Z.
Int J Health Geogr. 2009 May 12;8:27.

Response: Thank you for your suggestion. We added the reference after sentence “Long-term exposure to PM2.5 is associated with shorter life expectancy and higher mortality risks from lung cancer and cardiovascular diseases (5-8).”

What about the role of indoor cooking on lung cancer risk?

Household air pollution and lung cancer in China: a review of studies in Xuanwei.


Home kitchen ventilation, cooking fuels, and lung cancer risk in a prospective cohort of never smoking women in Shanghai, China.
Cancer risk from gaseous carbonyl compounds in indoor environment generated from household coal combustion in Xuanwei, China.


Response: Thank you for your suggestion to consider indoor combustion. We did discuss the role of indoor cooking and lung cancer. “Our estimates may be conservative since not all time-varying covariates were considered in our model, such as indoor biomass combustion (48-51). Although most countries included in this study were high-income countries and used a limited proportion of indoor biomass combustion, the true effect of coal power plants might be even higher if biomass combustion remained constant rather than decreasing. We adjusted total coal consumption in the model, which included the indoor combustion.” We modified the manuscript and added the above references into the manuscript. (new page 14, new paragraph 2, new line 9-10)

Additional sources of data on tobacco smoking trends.

http://gamapserver.who.int/gho/interactiveCharts/tobacco/use/atlas.html

WHO global report on trends in prevalence of tobacco smoking 2015

http://apps.who.int/iris/bitstream/handle/10665/156262/978924156492?sequence=1

Response: Thank you for your kind help to provide the data. Our data for smoking prevalence were from Ng M et al’s work, published on JAMA in 2014. The research has the earliest and exclusive data on smoking prevalence for almost every country (Ng M, et al, 2014). To consider the consistency of data, we stucked to the exclusive data including the smoking prevalence from 1980 to 2012.
Figures 2 and 3 are impossible to read. Perhaps they should be separated into four each.

Response: Thank you for your kind reminder. We did submit figures with very high resolution. Unfortunately, the files might be compressed before presenting in front of you. We re-provided enlarged figure 2 and figure 3 to review and re-submitted higher resolution figures online as well.

p. 14
Our identified confounders associated with both coal capacity and lung cancer at the 17 national level included adjustments for the appropriate latency period and strong temporality 18 justifications for causal inference (41).

"Univariate, behavior-environmental, 5-year-lag, 10-year-lag and 15-year-lag models were 19 applied to examine the effect among males and females, respectively (Table 2). The point estimates 20 of per capita coal capacity among the year-lag models were similar, so we picked the 10-year-lag 21 model as our primary model."

Comment: Please discuss latency in greater detail. It is my understanding that lung cancer develops after smoking for several decades. Suggest redoing the analysis with the assumptions of 20- and 30-yr lag and seeing whether the correlations change.

"The model includes a three- to four-decade lag between a rise in the prevalence of smoking and a rise in smoking-attributable mortality" in Trends in smoking and lung cancer mortality in Japan, by birth cohort, 1949-2010. Funatogawa I, Funatogawa T, Yano E.
Response: Thank you for the chance to redo the analysis and discuss latency in details. We agreed that there are long time lags (20-40 years) for lung cancer mortality associated with smoking (Loeb LA, et al 1984) and we redid the analysis with the 20- and 30-year lag of smoking. The results remain similar and the smoking is still a strong risk factor to lung cancer. The effect of coal capacity on lung cancer incidence is still positively associated, but less significant. For our study period of 2000 to 2016, a 30-year lag of smoking infers data required from 1970, which is not available at global scale. The weaker significance of exposure of interest might be partially due to substantial missing data of many covariates in the early years. We put the redone analysis in supplementary table 3 as sensitivity analysis and added the following sentence, reading: “Longer lag time of smoking of 20 and 30 years were also applied as sensitivity analysis (Supplementary Table 3).” (new page 9, new paragraph 4, new line 21-22 and supplementary table 3) in results and “There is little difference between the lag 5 and lag 10 models in terms of quadratic information criterion (QIC) (12) and coefficients, and longer period of latency for smoking also yields similar results. Therefore, for sake of consistency with the other covariates, we fix lag 10 for coal capacity as primary model and estimate PAFs.” in discussion. We also added the reference as you suggested.” (new page 12, new paragraph 1, new line 4-5)

Also, if Hill's criteria are to be invoked (Ref. 41), it would be useful to list the criteria considered and indicate how they are or are not satisfied.

Response:

Hill put nine criteria to diagnose causality in 1965. There are:

(1) The strength of association.

(2) Consistency from the different studies.

(3) Specificity for the strong association between diseases and workers.

(4) Temporality of cause and effect.
(5) Biological gradient: dose response curve.

(6) Plausibility: the causation we suspect is biologically plausible.

(7) Coherence: the causation we suspect is not conflicted with general known facts.

(8) Experiment: Is the observation violating experimental evidence?

(9) Analogy: Do we see slighter different but similar effect from similar circumstance?

The statistical association between coal-fired power plant and lung cancer in the study is strong (criteria 1 and criteria 5), after adjusting possible confounders and different time lags (criteria 4). Air pollutant (PM2.5), heavy mental, and other hazardous volatile organic compounds from coal-fired power plants are well known to be associated lung cancer (criteria 6, 7, 8). There are some researches also discussing the possible effect from coal-fired power plants (criteria 9 and 3). In conclusion, we don’t find any serious violation against Hill’s criteria and the causality between coal-fired power plants and lung cancer is likely.

Significant digits. The general rule is that no more non-zero digits should be given than are justified by the uncertainty of the value.

https://www.hccfl.edu/media/43516/sigfigs.pdf

If the uncertainty is greater than about 7%, only two non-zero digits are justified.

Thus, for example, in Table 1

GDP (PPP) 742.85 (573.38~912.31)

Should be 740 (570-910)
Response: Thanks for your correction. Our principle is to keep significant digits (2 digits after decimal point) consistent with our data and across the whole article as most scientific papers do. For table 2, since many 95% confidence intervals of some covariates are relatively small, taking 3 significant digits will make the intervals round down to 1.00. Readers might not be able to understand the original number is below or above 1.00. Therefore, we initially picked up 3 digits after decimal point to report data in table 2.

We modified the significant digit as your suggestion to justify the uncertainty of values. We would also like to keep consistency of significant digits through whole paper at the same time. We adjusted to 2 digits after decimal point as you suggested below in the tables and results accordingly.

Smoking prevalence c
Males 32.23 (31.16~33.3)

Comment: The small range of the 95% CI does not make sense. Smoking rates vary considerably by country. Same comment for traffic index, industrialization index.

Response: Thank you for giving us a chance to explain the issue you brought up. We initially reported confidence intervals (CIs) to represent the index on the global level. That is, we applied standard errors (=standard deviation/√N) of the data, rather than standard deviation (SD) to estimate CIs of true population. We agree that it is better to present the characteristics of the data on the population level, and therefore we presented quantiles instead. We modified results on the manuscript and Table 1. (new page 9, new paragraph 1, new line 1-2)

The increased per capita coal capacity is associated with the higher risk of lung cancer by a factor of 5.68% (=1.590.12) while the decreasing smoking prevalence prevented the risk by a factor of 11.28% (=1.033.50). (new page 12, new paragraph 3, new line 23 to new page 13, new paragraph 1, new line 1)

In Table 2, RR should be given to two decimal places unless the 95% CI values are within 0.02 of the RR value.
Response: Thank you for your reminder. We modified according and took 2 digits after decimal.”

Please review numbers in the text and tables and adjust accordingly.

Response: Thank you for your kind reminder. review numbers in the text and tables again and adjust accordingly as mentioned in above.

Reference:


Reviewer #2: 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.

Response: We would like to thank you for reviewing our manuscript and providing constructive suggestions, which help improve the quality of this manuscript. We have carefully revised the manuscript and our responses are provided as below.
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.

Response: Thank you for an opportunity to explain ecological fallacy. In 1950, Robinson published a paper showing the contradictory of the outcomes from ecological studies and from individuals (Robinson, 1950). Many scholars started aware of the issue and many public health practitioners seem afraid of ecological studies. In 2011, epidemiologist Idrovo explained clearly where ecological fallacy would present and proposed three criteria to diagnose ecological fallacy (Idrovo, 2011), they are:

(1) Results must be obtained with ecological (population) data

(2) Data must be inferred to individuals.

(3) Results obtained with individual data are contradictory.

All three of these should be present to confirm the existence of ecological fallacy.

The data collection in our study was from population data, however, we have no intention to infer the results to individuals (criteria 2). In addition, the results we have were very comparable to most current studies (criteria 3) as discussed in Discussion. Therefore, ecological bias should be minimal in the study.

However, residual and unmeasured confounders, such as national-level educational attainment or occupational exposures, may exist and bias results, no matter it’s an ecological study or not; adding more parameters to our analysis would destabilize estimates and cause loss of statistical power. Other unmeasured meteorological factor such as wind directions, and/or geographical factors, cannot be adjusted in our model and might act as confounders. However, since neither the electricity matrix nor meteorological/geographical factor is relevant to a country’s healthcare system, misclassification is non-differential and more likely biases toward the null. We modified the following sentence in limitation, reading: “However, residual and unmeasured confounders, such as national-level educational attainment or occupational exposure, may exist; adding more
parameters to our analysis would destabilize estimates and cause loss of statistical power. Potential misclassifications of meteorological factor such as wind directions, and/or geographical factors, cannot be adjusted in our model. Since neither the electricity matrix nor meteorological/geographical factor is relevant to a country’s healthcare system, misclassification is non-differential and more likely biases toward the null.” (new page 13, new paragraph 3, new line 20-23 to new page 14, new paragraph 1, new line 1-3)

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.

Response: Thanks for your comment and we fully agree that not just PM but other carcinogenic compounds are produced during coal combustion. Heavy metal, such as mercury is also a well-known by-product from coal-fired power plants (Liu, 2018) and associated with lung cancer. (Ellingen et al, 1993). Therefore, instead of using SOx, NOx, PM or any component of air pollutants, we picked up sizes of coal-fired power plants (MW) and investigate the association between the two at national level. Therefore, the health effects of different pollutants from coal combustion are already taken into consideration in the study design. We address your concern in Discussion, reading: “Since all pollutants related to lung cancer are not known, and known pollutants compose a small fraction of PM2.5, per capita coal capacity could serve as a better estimate of externality then pollutant composition measurements. Those pollutants such as SOx, NOx, heavy metal are associated with lung cancer from previous studies (34).” (new page 11, new paragraph 2, new line 13-14)

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.

Response: Thanks for your suggestion. We revised the abstract accordingly, reading: “These analyses suggest an association between lung cancer incidence and increased reliance on coal for energy generation.” (new page 2, new paragraph 4, new line 22-23)
3) Introduction, page 4, rows 7-9: particulate matter is associated with cardiovascular risk also in subjects without cancer.

Response: Thanks for your kind reminder. We revised the manuscript as following sentence: “Long-term exposure to PM2.5 is associated with shorter life expectancy and higher mortality risks from lung cancer and cardiovascular diseases (5-8).” (new page 3, new paragraph 1, new line 7-8)

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).

Response: Thanks for your suggestion. We modified the sentence accordingly as: “assuming no residual confounding or other sources of bias, GEE produces unbiased estimates of the beta coefficients, regardless of the within-country correlation structure specified, although a specification closer to the true correlation structure leads to lower standard errors” (new page 6, new paragraph 2, new line 10-13)

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.

Response: Thanks for your suggestion. We revised accordingly as following “and C18 to C21 in ICD-10” and (new page 7, new paragraph 1, new line 2)

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

Response: Thanks for your correction and we revised accordingly. “We estimate the population attributable fraction (PAF) of lung cancer to coal-fired power plants in 2015 and predict the PAF in 2025 among studied countries.” (new page 7, new paragraph 2, new line 5)

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.

Response: Thank you for giving us a chance to explain. The incidence rate difference is 45.68 per hundred thousand from first period (454.07) to the last period (408.39). We modified the manuscript to make it clear as following: “From the first period to the last, average age-
standardized incidence rates from lung cancer decreased by 45.68 (i.e., from 454.07 to 408.39) per hundred thousand (10.06%) in males but increased by 11.36 (i.e., from 143.50 to 154.86) per hundred thousand (7.92%) in females.” (new page 8, new paragraph 2, new line 18-19)

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.

Response: Thank you for your question and suggestion. Another reviewer, however, liked to keep the table and he strongly suspected the variation of many covariates are too small. We do believe that providing demographics of studied subjects, including population helps our readers to better understand the characteristics of included subjects we inferred to. Instead of reporting 95%CI, we reported 2.5th and 97.5th quantile of covariates in selected countries since some covariates are not normally distributed (Table 1). (new page 9, new paragraph 1, new line 1)

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).

Response: Thank you for your kind correction. We double checked the data and correct the manuscript accordingly as following: “With a 1 KW increase of coal capacity per person in a country, the relative risk of lung cancer increases by a factor of (95%CI=7%~1.35%) among males and 85% (95%CI=22%~182%) among females.” (new page 10, new paragraph 1, new line 3-4)

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

Response: Thank you for your suggestion. We agreed that “adjusted for different variables in different models” is redundant since the table is self-explanatory. We modified the title as “Table 2. Relative risk (RR) and 95% confidence intervals (CIs) of the increase in lung cancer incidence with change in coal capacity, among males and females.” (new page 10, new paragraph 2, new line 8-9)

11) Supplemental Table 3: please check for the correctness of the "RR" empty columns.
Response: Thank you for your careful check. We deleted the column since the “RR” has been estimated in Table 2. The revised supplementary table is numbered 5. (new page 10, new paragraph 4, new line 16)

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.

Response: Thank you for your comment. We did include smoking prevalence to adjust the possible confounding since tobacco smoking might be collinear with other socioeconomic conditions as you mentioned. We added your suggestion and modified the sentence, reading: “we are still interested in considering the nuanced differences of smoking prevalence and included in the model since it might be collinear with uncontrolled confounding from occupational exposures.” (new page 14, new paragraph 2, new line 19-21)

13) Please check the correctness of the authorship in citation 26 (Liang & Zeger).


References:

