Reviewer's report

Title: Body mass index and lung cancer risk: A pooled analysis based on nested case-control studies from four cohort studies

Version: 0 Date: 31 Jul 2017

Reviewer: Traci Bethea

Reviewer's report:

This manuscript is an interesting approach to the study question about the association between BMI and lung cancer risk and benefits greatly from a nested case-control design and a large study population.

Abstract: Since the abstract does not present data for BMI<18.5, please present a p-interaction that excludes these participants.

Introduction: Please clarify high BMI or BMI>=30 kg/m2 as the proxy measure for obesity that has been inversely associated with lung cancer risk. Since Mendelian randomization is not a component of the present study, it is unclear why it is a focus of the introduction.

Materials and Methods: How do histologically confirmed cases differ from cases confirmed through cancer registries, review of medical records, or diagnosed clinically? Were controls cancer-free at baseline for matching or did controls have to remain cancer-free over follow-up? How were cases and controls matched on age (for example, exact age, plus/minus 5 years, frequency matching)? The matching criteria for CARET is substantially different than for the other studies. What is the rationale for different matching strategies across studies? Can the analysis be repeated with cases and controls from CARET matched using a more similar approach? Was histologic type included in the information requested from each study or were these data acquired another way? Were other potential covariates considered? Were the portions of EPIC with self-reported weight and height excluded from the sensitivity analysis? If not, these participants should be removed and the sensitivity analysis should be repeated. Since pack-years is the variable for adjustment by smoking, did the authors consider stratifying by pack-years instead of smoking status? For the test for heterogeneity, how was BMI modeled? If BMI was modeled as a continuous variable, given that the study demonstrates non-linearity in the association, the authors should consider another approach to modeling the exposure of interest. How was the reference group for the RCS function selected? How were tests for interaction and tests for linearity conducted?

Results: Was a statistical interaction with gender present? The paragraph that describes the findings for Figure 1 should be presented before the paragraph that describes the findings for
Table 5. Please add sample results or add a supplementary table to present ORs for lung cancer diagnosed within the first 3 years of follow-up and lung cancer diagnosed after 3 years of follow-up. Although the p-heterogeneity is <0.05, the I² is approximately 50%. This magnitude of variation due to heterogeneity is of concern even without statistical significance. The authors should address the heterogeneity between studies.

Discussion: The discussion relies heavily on 2 meta-analyses and may include too many numerical findings. It would be stronger with additional development to place the current findings in the context of the rest of the literature. A strength of the study is the lack of recall bias due to prospective data collection on BMI. It is unclear what is meant by "comprehensive information on BMI," so the authors may want to clarify or remove this phrase.

Table 1: Column 3 uses "enrollment" while column 5 uses "enrolment." The authors should use a consistent spelling throughout. Also, column 7 could be named "Source of height and weight data." For EPIC, the authors may have intended to use "except" rather than "expect."

Table 2: The authors may mean "missing" rather than "ever" for smoking status. Please add the distributions of BMI and histologic subtype to the table. Since time elapsed has been defined for both cases and controls, please add the distribution of the data for controls. Age at diagnosis is an input for time elapsed but does not factor into the analyses, so it is unclear why it is included in the table.

Tables 3 and 4: These tables present data in a similar manner and could be combined. Please clarify that pack-years is the variable being used to adjust rather than smoking. In Table 4, please relocate the subscripts to the stratum headings so that it is clear which stratum-specific analyses include which covariates.

Table 5: Given the range in diagnosis years, one would expect that there were unspecified (nos) tumors. Were these tumors and carcinoid tumors excluded from the analyses? If not, please add these strata to the table or note the sample sizes in a footnote.
Are the methods appropriate and well described?
If not, please specify what is required in your comments to the authors.

No

Does the work include the necessary controls?
If not, please specify which controls are required in your comments to the authors.

Yes

Are the conclusions drawn adequately supported by the data shown?
If not, please explain in your comments to the authors.

Yes

Are you able to assess any statistics in the manuscript or would you recommend an additional statistical review?
If an additional statistical review is recommended, please specify what aspects require further assessment in your comments to the editors.

I am able to assess the statistics

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

Needs some language corrections before being published

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 organization 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.