Reviewer’s report

Title: A cluster analysis of chronic obstructive pulmonary disease in dusty areas cohort identified three subgroups

Version: 0 Date: 06 Jul 2017

Reviewer: Laura Paulin

Reviewer’s report:

In their paper "A cluster analysis of chronic obstructive pulmonary disease in a dusty area cohort identified three subgroups," Kim et al describe their findings of a principal component analysis done in a cohort of 272 patients with COPD. The authors describe 3 subgroups, 2 which have milder disease, and describe the changes in clinical outcomes over one year of follow up by each subgroup. There is a great deal of interest in identifying phenotypes that may experience worse outcomes in COPD, and this study is likely of interest to a general readership. However, there are several points that need addressing prior to potential publication. In particular, the methods could be more descriptive and the results more comprehensive. Please see specific comments below.

Methods:

* "Urinalysis" was included as a variable in the factor analysis. Can the authors expand on what component of the urinalysis was included in the analysis? (RBC, nitrates, WBC, etc).

* Would expand the methods to include more information about the CT scan procedure. There needs to be a better a description of what the emphysema index is, and also describe the airway wall thickness measurement since this is addressed in the discussion.

Results:

* Page 8, line 6. The authors state that "Subgroup 2 patients experienced…the most rapid decline of FEV1." May want to clarify in the text that this was not a statistically significant difference in FEV1 decline compared to the other groups.

* Page 8: The authors state that subgroup 2 had the most frequent exacerbations at 15.2%, however would clarify in the text that this is not significantly different that subgroup 3 (14.8%).

* Page 9: "Subgroup 2…had the most frequent exacerbations--" is this true?? Was there a difference between group 2 and group 3? Would clarify in the text.

Discussion:
Page 9: prior research suggests that biomass exposure may be associated with phenotypic differences, but the authors state that in their study this did not appear to be the case since the three subgroups had a similar biomass exposure. Are the authors confident in this conclusion and of the assessment of biomass exposure? How was the biomass exposure measured? Was this a validated questionnaire? The authors should explore the possibility of residual confounding and perhaps add this as a limitation.

The authors explain in their limitations that duration and intensity of exposure to cement dust is unknown, and comment that there may be an interaction between other environmental exposures (including cigarette smoke) and cement dust, which this study is unable to comment on. However, in the discussion, the authors summarize that the current study "is a unique study sample for demonstrating environmental exposure to cement dust." It may be misleading to state that this study was designed to show how environmental exposures may influence COPD given the lack of exposure assessment/characterization. To be able to demonstrate that this cohort had meaningful environmental exposures, would want to know how was exposure measured, was there a gradient of exposure estimated, etc. If data to address the exposure characterization is unavailable, would modify the discussion to avoid saying that this is a cohort designed to look at environmental exposures.

As the authors suggest that this cohort may have significant environmental exposures, how may these subgroups vary in a population that does not live in proximity to a cement plant?

There are several references to the airway wall thickness measurement in the discussion. If this is an important outcome, would highlight this is in the methods and results. The description of how the mean airway wall thickness was obtained was not included in the methods and the results are somewhat hidden in Table 2 and not highlighted in the results section.

Would add a limitation about the relatively small sample size of this cohort in addition to location and disease severity limitations.

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.

No

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 recommend additional statistical review

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