Reviewer’s report

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

Version: 0 Date: 24 Jul 2017

Reviewer: Gregory Kinney

Reviewer’s report:

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

This manuscript reports on an analysis of 203 members of the CODA cohort in Korea.

Major Comments

Major 1: The statistical methods employed for this analysis are unclear. The variables included in the PCA are stated and the PCA method is described but the authors then used the PCA results to perform a clustering analysis and that aspect is not well described. Stating "Hierarchical cluster analysis using Wards method was performed with principle factors extracted from factor analysis" is not sufficient for the reader to understand the methodology.

Major 2: this comment also relates to the PCA analysis, specifically the selection criteria used to choose the variables included. [A] The reader would benefit from the authors showing the correlational structure of the variables before the PCA analysis was performed. [B] The reader would benefit from knowing how the variables were transformed in order to be included in the PCA as highly variable variables will drive factors. [C] the authors report a KMO statistic of 0.593 which indicates a poor measure of sampling adequacy. The Bartlett's test is significant but the ratio of rows (203) to variables (15) is not less than 5 (13.8) indicating a tendency for the test to be significant even when the variable selection for the PCA may not be appropriate. Also four of the PCA factors have less than three variables loading strongly suggesting that the factors may not be reliable. The authors should consider performing a KMO index calculation for each variable in order to assess their inclusion. This reference briefly discusses that approach, how to interpret the results and how to alter the PCA itself if necessary https://eric.univ-lyon2.fr/~ricco/tanagra/fichiers/en_Tanagra_KMO_Bartlett.pdf

Major 3: the delta outcomes of interest to the analysis, FEV1, mMRC and CAT score are included in the PCA, the authors should discuss how this might affect the association between factors that include each and how they relate to the change in those variables over time. I don't believe that it's inappropriate to look at change in variables included in a PCA derived factor but the change variable is not independent of the PCA in this case and the reader would benefit from considering this.

Minor Comments
1) There are minor language usage edits that the authors should consider for clarity, for instance page 4, line 21 "The initial questionnaire included the data of demographics, disease history..." would read better as "The initial questionnaire included demographics, disease history...". This is not the only example and the authors should consider further editorial steps.

2) In table 1 the first and second row label are inverted, Eigenvalues are not typically in the range of percent variance explained (also the label for "% variance" should be "% variance explained").

3) The footers for each table and figure should include the statistical test resulting in the p value reported so that the table and figure can stand alone without reference back to the methods sections for clarity for the reader.

**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.

Unable to assess

**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'm working on a manuscript (first author) using very similar methodologies in a very similar disease right now. I am comfortable ethically that this work is not influencing my review of the current manuscript.

Other than 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.