Reviewer’s report

Title: Distinctive subgroups derived by cluster analysis based on pain and psychological symptoms in Swedish older adults with chronic pain - a population study (PainS65+)

Version: 0 Date: 25 Jul 2016

Reviewer: Victoria Strauss

Reviewer's report:

It is generally well written and easy to follow when reading it through; however, I do have major concerns as described below.

1. The authors used the two-step cluster analysis to identify subgroups of people with chronic pain and psychological symptoms. Given that the choice of methods is crucial to the paper, I wonder whether the authors explored different choices of clustering algorithm? Do different algorithms lead to similar results? What do the authors mean "TSCA has several proper features that distinguish it from traditional clustering techniques (p9 in Statistics section)"? Is the two-step cluster analysis a combination of k-mean and hierarchical?

2. Has the authors considered model based clustering, latent class analysis? My main concern about the two step cluster analysis is that the number of clusters was pre-defined based on "best fit of hierarchical methods". Does it mean that there is no statistical way to test whether an additional cluster is needed or not? The two step cluster analysis also does not evaluate how each individual fit to profiles of the subgroup an individual is assigned to.

3. What are the assumption of variables included in the two step cluster analysis? Do they need to be Normal distributed? If yes, is there any transformation done before clustering is applied?

4. Can the authors provide a flow chart to describe the number of people invited to the study, the number of people included in the analysis, the reasons for exclusions and the number of people included in the final analysis? Can they compare baseline characteristics of people included with those excluded in the analysis?

5. They stated that 4154 people were excluded due to their criteria of chronic pain (p10). Can the author explain their criteria for chronic pain?

6. Can they authors compare baseline characteristics between people excluded due to missing data and those included in the analysis?

7. It stated on p10 that 9clusters were suggested by BIC, but, the authors decided four clusters are the optimal number. Since one advantage of the two step cluster analysis vs the
traditional k mean cluster analysis is that it has procedures to select the number of clusters needed, I wonder whether they decided to overrule the BIC suggestion. Have the authors explore other models with a different number of clusters ranging from 5 to 9 as suggested by BIC?

8. I like the idea of identifying subgroups of people with chronic pain. However, what are the clinical implications of this study? How can the results be used in clinical practice?

**Are the methods appropriate and well described?**
If not, please specify what is required in your comments to the authors.

Unable to assess

**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:

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 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