Reviewer’s report

Title: Identifying heterogeneous health profiles of primary care utilizers and their differential healthcare utilization and mortality – a retrospective cohort study

Version: 0 Date: 10 Dec 2018

Reviewer: Patricia Kipnis

Reviewer's report:

Title
Utilization should not be plural

Abstract
Methods section is confusing, specifically the sentence:

"We used latent class analysis to segment the population into k classes (k from 2 successively to 6) by disease"

This can be interpreted as first dividing the population into diseases and then segmenting each disease into 6 segments. This does not appear to be the method you used. If it was then you need to describe how the disease categories were defined and how patients were classified into difference diseases.

Methods
Study Design
What data are available in this system?
What are the diseases included in the Singapore chronic disease management program? How are they defined?
Did you include all the diseases in the Charlson and Elixhauser scores? How did you handle overlaps?

Model Selection
What was the highest k tested? K=6? Why?
How was clinical interpretability assessed?
What predictors/covariates were used in the analyses?
How is substantial size defined?
How would interventions be mapped to each segment?
How did you assess class homogeneity? The original aim of the study is to define more homogeneous clusters - how do we know the resulting classes are homogeneous? Would it be useful to further segment some of the larger segments you defined? Class 2? Class 5?

Statistical Analysis
Were the analyses run for each k? Where are those results shown?
How are "chronic disease patterns" defined?
How is healthcare utilization defined? Utilization information is only shown in the tables and the tables do not have any description of how the various visit and admission counts were defined (e.g. do you include only clinic visits? Telephone visits included? Do you include only inpatient admissions? Do you include outpatient surgeries taking place in the hospital setting? Observations? How are specialty visits defined? Where does mental health fall?)

Where patients who died at the end of 201 included in the LDA?

The description of the multivariate model on page 8 lines 182-185 is confusing. You describe a bivariate analysis and then jump into discussing a model without stating that a multivariate analysis was also conducted.

Clarify that the offset was included for the utilization outcomes (all?)

Line 193 - "determine significance" of what??

Line 197 - "number of days from entry to death" entry into what??

Results
The 100,747 patients were all alive at the end of 2012?
Can you color code the columns of Table 1? It is hard to discern any patterns? Or can you find a way to display the data so that one can discern the difference in patterns across the classes?
Add titles to the columns in subsequent pages of the tables - it is hard to follow once the table breaks.
Can you offer a description of the patients in the various classes? What diseases do they tend to have? Are they truly homogeneous?

Discussion
Again, a description of the type of patients included in each class would be helpful. Also, discuss where mental health falls and why it is not important to single it out.

Tables and Figures
See comment re: Table 1 above
Add footnotes to the tables describing the metrics

Table 2 - add references for the metrics used as footnotes. Add a column with class sizes
Were there other k>6 tested?

Table 3
Add a footnote describing the test used to arrive at the p-value

Table 4
What are the a and b subscripts for?
Define IRR in a footnote
Perhaps forest plots would be better than a table to describe the results as one can see the overlap across classes

Figure 1 is difficult to follow specially without color. Can you think of a different way to display the results to describe the disease distribution across classes?
Are the methods appropriate and well described?
If not, please specify what is required in your comments to the authors.

Yes

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 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