Reviewer's report

Title: The comorbidity burden of type-2 diabetes mellitus: patterns, clusters and predictions from a large English primary care cohort

Version: 0 Date: 30 Apr 2019

Reviewer: Amaia Calderon Larrañaga

Reviewer's report:

The present study aims to (1) quantify the comorbidity burden in people with T2DM, (2) estimate the prevalence of specific comorbidities, and (3) identify clusters of similar conditions. The main added value of the paper resides in the size and quality of the employed data source, which enables identifying a considerable number of incident cases of T2DM across a period of 10 years. The analyses are well performed and the paper is clearly written. There are, however, a few issues that require further clarification:

1. My main concern is related to the third objective of the study mentioned above, since I do not see how it contributes to improve our understanding of the comorbidity burden of T2DM. By applying a hard clustering algorithm such as the hierarchical clustering, we force each disease to be included in one single cluster, which is little compatible with the fact that clusters tend to overlap (and even more so with the pass of time, as people get older). In other words, the technique used in this study prevents the authors from obtaining a real picture of the disease clusters in people affected by T2DM at the time of diagnosis and during follow-up, when such clusters tend to become more complex in terms of the number and type of conditions (Prados-Torres A, PLoS One, 2012). In fact, currently, the take-home message related to the third objective is fairly weak.

2. Longitudinal data are used in two ways; (a) to study how the burden and type of comorbidity of all incident T2DM cases evolve over time and, (b) to analyse the trends of comorbidity burden in incident T2DM cases diagnosed yearly between 2007 and 2017. The description of the methods as well as the results related to these two types of analyses is something that needs to be improved and made clearer.

3. In relation to analysis (a) (as described above), the authors find that, during follow-up, the prevalence of depression and asthma decreased in all groups, while the prevalence of all other conditions increased. Some explanation in relation to this unexpected finding for depression and asthma would be appreciated.
4. One of the main conclusions of the paper is the expected increase in the prevalence of depression among future incident T2DM cases. However, these projections could be overestimated because of the way depression was identified in the database. By looking at prescription of antidepressants, not only other types of mental health problems mentioned by the authors themselves (obsessive-compulsive or bipolar disorder) but also health problems such as insomnia and pain (often prescribed off-label) may have been incorporated into the depression group. Given the relevance of this finding, the authors should try to quantify the extent of false positives identified by this procedure.

Minor mistakes:

* Please re-write the sentence in lines 194-195 to clarify that what is written in brackets refers to the number of comorbidities at T2DM diagnosis.

* Numbers in lines 202-203 referring to the prevalence of CHD in males do not match those found in Figure 2.

* The sentence in lines 217-219 is unfinished.

* Please check and correct the labels at the bottom of Figure 1.

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.

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

**Statement on potential review bias**
Please complete a statement on potential review bias, considering the following questions:

1. Did you co-author any publication with an author of this manuscript in the last 5 years?

2. Are you currently or recently affiliated at the same institution as an author of this manuscript?

If you can answer no to all of the above, write 'I declare that I did not publish with these authors in the last 5 years and also meet the affiliation criteria”. If your reply is yes to any, please give details below.

I declare that I did not publish with these authors in the last 5 years and also meet the affiliation criteria

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/](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